

**Fakultät  
der  
Universität Duisburg-Essen**

**Aus den Kliniken Essen-Mitte, Innere Medizin V,  
Klinik für Naturheilkunde und Integrative Medizin**

**Die Qualität publizierter Experimente in der  
Grundlagenforschung zur Homöopathie:  
Entwicklung und Anwendung eines Kriterienkatalogs**

**Inaugural - Dissertation  
zur  
Erlangung des Doktorgrades der Naturwissenschaften in der Medizin  
durch die Medizinische Fakultät  
der Universität Duisburg-Essen**

**vorgelegt von  
Dipl.-Biol. Beate Stock-Schröer  
aus Saarbrücken**

**2010**

**Dekan: Herr Univ.-Prof. Dr. med. M. Forsting**

**1. Gutachter: Herr Prof. Dr. med. Gustav Dobos**

**2. Gutachter: Frau Priv.-Doz. Dr. rer. nat S. Moebus**

**3. Gutachter: Frau Priv.-Doz. Dr. rer. nat. G. Ulrich-Merzenich, Bonn**

**Tag der mündlichen Prüfung: 1. August 2011**

### **Publikationen aus dem Jahr 2009**

Stock-Schroer, B.; Albrecht, H.; Betti, L.; Dobos, G.; Endler, C.; Linde, K.; Ludtke, R.; Musial, F.; van Wijk, R.; Witt, C.; Baumgartner, S.:  
"Reporting Experiments in Homeopathic Basic Research--Description of the Checklist Development." Evid Based Complement Alternat Med 2009 Nov 1. [Epub ahead of print]

Stock-Schroer, B., Albrecht, H., Betti, L., Endler, P. C., Linde, K., Ludtke, R., Musial, F., van Wijk, R., Witt, C., Baumgartner, S. (2009):  
Reporting experiments in homeopathic basic research (REHBaR)--a detailed guideline for authors. Homeopathy 98 (4), 287-98.

<b>1</b>	<b>Einleitung .....</b>	<b>6</b>
<b>1.1</b>	<b>GRUNDLAGEN DER HOMÖOPATHIE .....</b>	<b>6</b>
<b>1.2</b>	<b>STATUS QUO DER GRUNDLAGENFORSCHUNG ZUR HOMÖOPATHIE .....</b>	<b>7</b>
1.2.1	GEGENSTAND DER UNTERSUCHUNG / FRAGESTELLUNGEN .....	7
1.2.2	AKTUELLER ÜBERBLICK ZU DEN FORSCHUNGSERGEBNISSEN.....	8
<b>1.3</b>	<b>QUALITÄT IN DER GRUNDLAGENFORSCHUNG ZUR HOMÖOPATHIE.....</b>	<b>10</b>
<b>2</b>	<b>Ziele der Arbeit .....</b>	<b>12</b>
<b>3</b>	<b>Umfrage zu ausgewählten Standards in der Grundlagenforschung</b>	<b>13</b>
<b>3.1</b>	<b>DER FRAGEBOGEN .....</b>	<b>13</b>
<b>3.2</b>	<b>AUSWAHL DER BEFRAGTEN INSTITUTE .....</b>	<b>13</b>
<b>3.3</b>	<b>ERGEBNISSE DER BEFRAGUNG .....</b>	<b>15</b>
3.3.1	VERBLINDUNG UND RANDOMISIERUNG .....	15
3.3.2	POSITIV- UND NEGATIVKONTROLLEN .....	16
3.3.3	REPLIKATION .....	16
3.3.4	STATISTISCHE BERATUNG.....	17
3.3.5	ANMERKUNGEN.....	17
<b>3.4</b>	<b>ZUSAMMENFASSENDE DISKUSSION.....</b>	<b>17</b>
<b>4</b>	<b>Erstellung des Kriterienkatalogs REHBaR (Reporting Experiments in Homeopathic Basic Research) .....</b>	<b>19</b>
<b>4.1</b>	<b>HINTERGRUND.....</b>	<b>19</b>
<b>4.2</b>	<b>DER DELPHI-PROZESS ALS KONSENSUMETHODE .....</b>	<b>20</b>
<b>4.3</b>	<b>DER KRITERIENKATALOG .....</b>	<b>25</b>
<b>4.4</b>	<b>DISKUSSION UND BEDEUTUNG DES KRITERIENKATALOGS .....</b>	<b>33</b>
<b>5</b>	<b>Evaluation von REHBaR als Bewertungsinstrument für die Qualität von Publikationen.....</b>	<b>36</b>
<b>5.1</b>	<b>HINTERGRUND.....</b>	<b>36</b>
<b>5.2</b>	<b>METHODIK.....</b>	<b>36</b>
5.2.1	AUSWAHL DER GUTACHTER .....	36
5.2.2	AUSWAHL DER PUBLIKATIONEN .....	37
5.2.3	PILOTSTUDIE.....	37
5.2.4	HAUPTSTUDIE .....	38
5.2.5	QUANTIFIZIERUNG DER ÜBEREINSTIMMUNG – KAPPA - WERTE K UND K <sub>N</sub> .....	38
<b>5.3</b>	<b>ERGEBNISSE DER EVALUATION – MESSUNG DER RELIABILITÄT .....</b>	<b>41</b>

<b>5.4</b>	<b>DISKUSSION DER EIGNUNG DES KRITERIENKATALOGS REHBAR.....</b>	<b>48</b>
5.4.1	EIGNUNG DER K – WERTE ALS MASS DER ÜBEREINSTIMMUNG .....	48
5.4.2	EIGNUNG DER GUTACHTER.....	49
5.4.3	EIGNUNG DES KRITERIENKATALOGS.....	49
<b>6</b>	<b>Zusammenfassung .....</b>	<b>52</b>
<b>7</b>	<b>Literaturverzeichnis.....</b>	<b>53</b>
<b>8</b>	<b>Anhang .....</b>	<b>60</b>
8.1	ABKÜRZUNGSVERZEICHNIS.....	60
8.2	FRAGEBOGEN ZU AUSGEWÄHLTEN STANDARDS IN DER LABORPRAXIS.....	61
8.3	ÜBERSICHT DER ANGESCHRIEBENEN INSTITUTE .....	63
8.4	FRAGEBOGEN FÜR DELPHI I.....	68
8.5	REHBAR: ERLÄUTERUNGSKATALOG MIT BEISPIELEN .....	73
8.6	REHBAR – KATALOG ZUR BEWERTUNG DER GÜTE EINER PUBLIKATION ..	84
8.7	ÜBERSICHT DER BEWERTETEN PUBLIKATIONEN.....	87
<b>9</b>	<b>Danksagung .....</b>	<b>90</b>
<b>10</b>	<b>Lebenslauf .....</b>	<b>91</b>

# 1 Einleitung

## 1.1 Grundlagen der Homöopathie

Die Homöopathie gibt immer wieder Anlass zu kontroversen Diskussionen. In vielen Bereichen der Medizin wird die Homöopathie als unwissenschaftliche Methode abgelehnt, in der Bevölkerung steigt das Interesse dagegen. Laut einer repräsentativen Befragung des Instituts für Demoskopie Allensbach aus dem Jahr 2009 verdoppelte sich die Zahl derer, die sich bewusst für eine homöopathische Behandlung entschieden haben, von 24% im Jahr 1975 auf 57% im Jahr 2009 (Allensbacher Archiv, IfD-Umfrage 10039, Juni 2009). Während sich demzufolge mehr als jeder zweite Patient in Deutschland eine homöopathische Behandlung wünscht, steht die Homöopathie bislang nicht auf dem Lehrplan im Medizinstudium.

Einen Beweis für ihre Wirkung oder eine Aufklärung des Wirkprinzips ist die Homöopathie trotz vieler Versuche bislang schuldig geblieben. Im Rahmen der klinischen Studien entbrennt immer wieder der Streit, ob die homöopathische Therapie nichts anderes als eine Behandlung mit Placebo ist. Eine Metaanalyse sämtlicher klinischer Studien zu Homöopathie versus Placebo aus dem Jahr 1997 (Linde et al. 1997) kam zu dem Ergebnis, dass Homöopathie nicht einer Placebowirkung zugerechnet werden kann. Eine erneute Analyse der Publikationen, ergänzt mit aktuellen Studien, fand hingegen bestätigt, dass die Homöopathie auf einen reinen Placeboeffekt zurück geführt werden kann (Shang et al. 2005). Aus biometrischer Sicht wurde allerdings kritisiert, dass durch die Auswahl und Berechnung der verwendeten Daten eine Verzerrung der Ergebnisse vorliegt (Ludtke und Rutten 2008).

Die Homöopathie ist eine Arzneimitteltherapie, deren Grundprinzipien von dem deutschen Arzt, Apotheker und Chemiker Samuel Hahnemann (1755 - 1843) entwickelt wurden. Dabei beruht die Wirkung der Therapie aus Sicht der homöopathischen Lehre auf vier Grundgedanken:

1. „Similia similibus curentur“ (lat.: „Ähnliches möge mit Ähnlichem behandelt werden“)  
Arzneimittel werden nach dem Ähnlichkeitsprinzip verordnet. Es werden solche Substanzen verwendet, die beim Gesunden ähnliche Symptome hervorrufen, die bei der zu behandelnden Krankheit auftreten.

2. Arzneimittelprüfung

Um das Simile-Prinzip anwenden zu können, muss bekannt sein, welche Symptome die Arzneistoffe bei Gesunden hervorrufen. Deswegen werden diese Arzneimittelprüfungen (HAMP) an gesunden, freiwilligen Personen durchgeführt.

3. Individualisierungsprinzip

Neben einer Bestandsaufnahme der Gesamtheit der Symptome werden in der homöopathischen

Anamnese Modalitäten, erstmaliges Auftreten bestimmter Symptome und weitere persönliche Begebenheiten erfragt. In der anschließenden Repertorisation führen alle diese Informationen durch die Gewichtung der einzelnen Symptome zu einer individuellen Mittelfindung.

#### 4. Herstellung (Potenzierungsverfahren) der Arzneimittel

Die in der Homöopathie verwendeten Arzneimittel werden schrittweise z.B. mit Äthanol verdünnt und bei jedem Verdünnungsschritt mechanisch verschüttelt oder weiter mit Milchzucker verrieben. Durch die mechanische Prozedur wird die Substanz nach homöopathischem Verständnis „dynamisiert“. Dabei existieren verschiedene Herstellungsverfahren, bei denen sich das Mischungsverhältnis von Substanz zu Lösungsmittel unterscheidet (in 1:10 Schritten = D-Potenzen, 1:100 = C-Potenzen oder 1:50.000 = Q- oder LM-Potenzen).

Die Homöopathie und vor allem die Verwendung hoch verdünnter Substanzen, bei denen eine nicht-molekulare Wirkung voraus gesetzt werden muss, ist immer wieder Gegenstand der Kritik. Nach den Berechnungen von Avogadro und Loschmidt kann in einer Lösung D23 bzw. C12, die um mehr als den Faktor  $10^{23}$  verdünnt wird, rein rechnerisch kein einziges Molekül der Ausgangssubstanz mehr enthalten sein. Homöopathische Zubereitungen, die mit potenzierten Verdünnungen von Stoffen jenseits der Avogadro-Konstante hergestellt werden ( $\geq$  C12, D23), werden als Hochpotenzen oder UHD (Ultra High Dilutions) bezeichnet. Niedrigpotenzen sind Homöopathika mit einer Konzentration kleiner als D12.

## 1.2 Status Quo der Grundlagenforschung zur Homöopathie

### 1.2.1 Gegenstand der Untersuchung / Fragestellungen

Die Grundlagenforschung zur Homöopathie befasst sich wie die klinische Forschung mit der Frage nach der empirischen Evidenz der Homöopathie. Darüber hinaus ist das zugrunde liegende Wirkprinzip Gegenstand der Untersuchungen. Dabei stehen drei Fragestellungen im Fokus der Forschung: 1. Unterscheidet sich eine potenzierte Substanz nur vom Lösungsmittel oder auch von der verdünnten Substanz in ihrer Wirkung? 2. Zeigen verschiedene Potenzen eine jeweils spezifische biologische Aktivität? 3. Folgt die Wirkung dem Simile-Prinzip? Neben den Experimenten zur möglichen Wirkweise verfolgt die Forschung auch das Ziel Informationen zum Thema Arzneimittelsicherheit und Systemstabilität zusammenzutragen: z.B. ideale Produktions- und Lagerbedingungen, inerte Aufbewahrungsmaterialien, optimierte Vorgehensweisen beim Verdünnen und Verschütteln (Baumgartner 2009). Welchen Einfluss das Material des jeweiligen Aufbewahrungsgefäßes auf die homöopathische Lösung haben kann, hat Claudia Witt in ihrer Doktorarbeit untersucht (Witt 2000). Eine weitere Studie bestätigt, dass Ionen aus den

verwendeten Aufbewahrungsmaterialien in die flüssigen Homöopathika übergehen und so die Wirkweise beeinflussen können (Ives et al. 2010).

Im Jahr 2006 waren über 900 Publikationen mit mehr als 1.100 Experimenten zur Grundlagenforschung in der Homöopathie registriert (van Wijk und Albrecht 2007). Generell lässt sich die Grundlagenforschung zur Homöopathie in vier Bereiche einteilen: Tierversuche, Pflanzenversuche, in-vitro-Forschung und physiko-chemische Experimente (Baumgartner et al. 2008).

Folgende Fragen stehen im Zentrum der Forschung: Verhalten sich komplexe Systeme anders als isolierte Gewebestrukturen und Zellen? Wie unterscheiden sich die homöopathischen Substanzen bei gesunden und kranken Untersuchungsmodellen? Welche Rolle spielt der genaue Herstellungsprozess, Verschüttelungs- und Verdünnungsvorgang? Welche Kontrollsituation bzw. Kontrollsubstanz ist für die jeweilige Fragestellung adäquat?

### 1.2.2 Aktueller Überblick zu den Forschungsergebnissen

Die neuesten Ergebnisse aus der Grundlagenforschung zur Homöopathie lassen sich wie folgt zusammenfassen:

1. Bislang hat keine externe, exakte **Replikation** (=Wiederholung eines wissenschaftlichen Versuches unter möglichst gleichen Versuchsbedingungen) zu einem identischen Ergebnis geführt (Vickers 1999; Baumgartner 2009). Replikationen in Italien und der Schweiz führten mit dem gleichen Arzneimittel bei dem gleichen Versuchsmodell zu einem signifikanten Effekt, allerdings fiel dieser gegensätzlich aus. In den Versuchsreihen in Bologna von Betti et al. konnte wiederholt eine Beschleunigung des Wachstums der vorgeschädigten Weizenkeimlinge mit der Potenzstufe D45 von Arsen statistisch signifikant nachgewiesen werden (Betti 1997). Bei den Replikationsversuchen in der Schweiz führte die gleiche Potenz des Arsens hingegen zu einer Wachstumshemmung (Lahnstein et al. 2009).

Es zeigt sich bei verschiedenen Modellen, dass Labor-intern die Ergebnisse repliziert werden können, auch wenn der Experimentator wechselt. Endler et al. verglichen alle Experimente, die mit homöopathischen Präparationen in der Konzentration  $10^{-23}$  an verschiedenen Modellen durchgeführt wurden (Endler et al. 2009). Sie kommen zu dem Ergebnis, dass von 107 Studien 77 Wiederholungsstudien waren, von denen 53 (69%) von vergleichbaren Ergebnissen berichteten. Dabei ist zu beachten, dass diese Experimente nicht als identische Wiederholungsversuche konzipiert waren und sich z.B. die verwendeten Zellen, Substanzen, Messmethoden und z.T. auch die Laborbedingungen voneinander unterschieden.

2. Im Bereich der **Pflanzenforschung**, die an gesunden und kranken Pflanzen Fragestellungen zur Homöopathie bearbeitet, liegen bislang zwei systematische Reviews vor. Die Autoren des



ersten Reviews kommen zu dem Ergebnis, dass bei den wenigen qualitativ guten Experimenten mit gesunden Pflanzen eindeutige Hinweise auf eine Wirkung sowohl von Niedrig- als auch Hochpotenzen zu finden sind (Majewsky et al. 2009). Der zweite Review gibt einen Überblick zu den Ergebnissen von homöopathischen Wirkungen an phytopathologischen Modellen. Da die in diesem Review gesichteten Publikationen überwiegend von schlechter Qualität waren, lassen sich hier keine gesicherten Erkenntnisse ableiten (Betti et al. 2009).

Bei Versuchen mit so genannten Potenzreihen (kontinuierlich aufsteigenden Potenzstufen) fällt auf, dass dicht neben einander liegende Potenzen z.T. gegensätzliche Wirkung zeigen: so kann eine D13 einen Effekt auf das Untersuchungsmodell haben, während die D14 keinen Unterschied zur potenzierten Wasserkontrolle zeigt (Scherr et al. 2009).

3. Die am häufigsten verwendeten **Tiere** in der Grundlagenforschung zur Homöopathie sind Ratten und Mäuse. In den letzten 10 Jahren liegt ein Forschungsschwerpunkt auf dem „Simile - Prinzip“. Bonamin und Endler kommen zu dem Ergebnis, dass gerade in höher entwickelten Organismen (komplexen Systemen) die homöopathischen Präparationen zu einer Stabilisierung des aus dem Gleichgewicht gebrachten Organismus führen (Bonamin und Endler 2010).

4. Ein weiteres Modell, das sowohl bei Pflanzen als auch bei Tieren untersucht wird, ist das **In-toxikationsmodell**. Die zugrunde liegende Theorie – die Isopathie (Gleiches möge Gleiches heilen) – besagt, dass nach einer Vergiftung z.B. mit Arsen dieselbe Substanz in homöopathischer Verdünnung zu einer schnelleren Wiederherstellung der Funktionen im Organismus führen kann. Linde et al. erfassten für ihre Metaanalyse 105 Publikationen (1955 bis 1993), von denen ca. 1/3 (31%) ausgewertet werden konnten. Bei diesen qualitativ hochwertigen Studien konnten deutlich positive Effekte der hoch verdünnten Substanzen im Sinne der Isopathie nachgewiesen werden (Linde et al. 1994).

5. Neben der in-vivo-Forschung gibt es zahlreiche **in-vitro**-Modelle, die z.B. mit humanen Zellen und Geweben durchgeführt werden. Während in vivo bedeutet, dass die Experimente mit komplexen, lebenden Organismen durchgeführt werden, werden als in-vitro-Forschung Versuche außerhalb eines lebenden Organismus in einer kontrollierten künstlichen Umgebung bezeichnet. Witt et al. analysierten für einen systematischen Review 75 Publikationen, die mit Hochpotenzen in vitro geforscht hatten. Davon konnten nur 18 Veröffentlichungen als qualitativ gut beurteilt werden, von denen 2/3 einen positiven Effekt der verwendeten Hochpotenzen zeigten (Witt et al. 2007).

6. Eine Übersicht aus dem Jahr 2000 verzeichnet 22 Publikationen zu **physiko-chemischen Experimenten**, die in der Zeit von 1966 bis 1998 mit überwiegend schlechter Qualität und geringer Aussagekraft veröffentlicht wurden (Witt 2000). In den letzten zehn Jahren sind aus diesem

Bereich neue Forschungsarbeiten publiziert worden. Mit Hilfe der UV-Spektroskopie haben Wolf et al. verschiedene homöopathische Aufbereitungen von Quarz, Kupfer und Sulfur bezüglich ihrer Struktur und damit Durchlässigkeit von UV-Strahlen untersucht. Sie folgern aus ihren Versuchen, dass die signifikant niedrigere Durchlassrate von UV-Licht in den homöopathischen Präparationen möglicherweise darauf zurückzuführen sei, dass homöopathische Lösungen weniger Strukturen aufweisen oder die Lösung an sich „dynamischer“ ist als die Kontrolle (Wolf et al. 2009).

### 1.3 Qualität in der Grundlagenforschung zur Homöopathie

Die Publikation ist das wichtigste Instrument, um Methodik, Ergebnisse sowie Diskussion der Studienergebnisse zu veröffentlichen. Eine detaillierte und adäquate Berichterstattung gehört zur Pflicht jedes Wissenschaftlers. Für die Grundlagenforschung in der Homöopathie gilt, dass die Publikation dem Leser eine kritische Beurteilung der Versuchsdurchführung und der Ergebnisse ermöglichen soll. Darüber hinaus soll sie alle Informationen enthalten, die nötig sind, um die beschriebenen Versuche ggf. im eigenen Labor wiederholen zu können. Ohne eine adäquate Beschreibung des Experimentes ist die Beurteilung der Validität und Reliabilität der Ergebnisse nicht möglich. In einem Bereich der Grundlagenforschung, der mit ultra hoch verdünnten Substanzen arbeitet und deren Ursache-Wirkungs-Beziehung unbekannt ist, muss Wert auf die Einhaltung wissenschaftlicher Standards gelegt werden. Die mangelnde Qualität in der Grundlagenforschung zur Homöopathie bzw. in der Berichterstattung der Ergebnisse wird von vielen Autoren aus den eigenen Reihen moniert( z.B. (Betti et al. 2009) (Witt et al. 2007)).

In den letzten zwanzig Jahren gab es einige Ansätze, Qualitätsstandards in der Grundlagenforschung zur Homöopathie einzuführen.

Baumgartner et al. formulierten als Einzige methodologische Standards u.a. zu Laborbedingungen, Wahl von adäquaten Kontrollen, Screening Tests sowie statistische Methoden (Baumgartner et al. 1998). Schulte publizierte einen Katalog, welche Angaben bei der Beschreibung von Experimenten im Bereich der Hochpotenzforschung erforderlich sind (Schulte 1994). Hornung et al. empfehlen für den Bereich der Homöopathie die Urtinktur (Ausgangssubstanz), den Verdünnungsprozess (Potenziervorgang), die Herstellung der Kontrollen sowie die Art der Anwendung zu beschreiben (Hornung und Linde 1991). Im Jahr 1994 wurde diese Liste auf insgesamt 23 Kriterien erweitert, die für Erstellung und Bewertung von Publikationen konfirmatorischer Experimente anwendbar sein soll (Linde 1994). (Konfirmatorische Experimente sind Versuche, die zur Verifizierung einer Annahme durchgeführt werden. Das bedeutet, für die Fragestellung lässt sich eine Hypothese formulieren, die es zu bestätigen oder zu widerlegen gilt. Ein exploratives Experiment wird durchgeführt, um eine Hypothese generieren zu können.) Die

Checkliste von Linde, die mit einem Punktescore arbeitet, wurde im gleichen Jahr für die Erstellung eines Reviews und einer Metaanalyse über Experimente aus der Toxikologie (Tierversuche) mit seriell erstellten Verdünnungen verwendet (Linde et al. 1994).

Im Bereich klinischer Studien wird Mitte der 1990er Jahre der Jadad Score (Moher et al. 1995) veröffentlicht. Mithilfe dieser siebenstufigen Skala soll es dem Leser ermöglicht werden, die Qualität einer Publikationen zu bewerten. Dabei betonen die Autoren, dass diese Skalierung nichts anderes als eine Richtgröße ist. Die Einteilung nach Zahlen zwischen eins und sieben diene dem Leser allein zur Abschätzung, wie wahrscheinlich es ist, dass die in der Publikation berichteten Methoden und Ergebnisse adäquat bzw. plausibel sind.

Witt entwickelte in ihrer Doktorarbeit (Witt 2000) zur physikalischen Untersuchung homöopathischer Hochpotenzen einen Bewertungsscore, der ähnlich wie der Jadad Score bei klinischen Studien ein sehr einfaches Punktesystem umzusetzen versucht. Dieses Bewertungsinstrument besteht aus acht Kriterien, eine Publikation kann maximal zehn Qualitätspunkte im sog. BEPEV Score erreichen. Dieser Score wurde danach modifiziert und als SAPEH Score für die Erstellung eines Reviews zur physiko-chemischen Grundlagenforschung in der Homöopathie angewendet (Becker-Witt et al. 2003) und nachfolgend für einen systematischen Review zur Beurteilung der Publikationen in der in-vitro-Forschung evaluiert (Witt et al. 2007).

Ein weiteres Bewertungssystem, das einen Punktescore verwendet, wurde im Jahr 2009 publiziert. Dieses wurde bislang für die Erstellung eines systematischen Reviews zu Experimenten mit gesunden Pflanzen (Majewsky et al. 2009) sowie für einen kritischen Review zur Wirkung von Homöopathika bei phytopathologischen Modellen (Betti et al. 2009) verwendet. Die Bewertung der Studien mit diesem Instrument erfolgt in zwei Schritten. Zum einen wird der MIS bestimmt: der „Manuscript Information Score“ mit einer maximalen Punktzahl von zehn Punkten in fünf Kategorien. Die zweite, nachfolgende Bewertung der Publikationen, die einen MIS  $\geq 5$  hatten, erfolgte über die SMEP „Study Methods Evaluation Procedure“. Hier sollte beurteilt werden, ob die publizierten Experimente verblindet und randomisiert waren und ob Positiv- und Negativkontrollen beschrieben wurden.

Zusammenfassend lässt sich feststellen, dass es bislang weder eine Richtlinie für die Erstellung von Publikationen noch einen konsensfähigen Katalog über Standards in der Grundlagenforschung zur Homöopathie gibt. Die vorliegende Arbeit soll einen Beitrag zur Qualitätssicherung darstellen und eine solche Richtlinie bzw. Katalog in einem Expertenteam entwickeln.

## 2 Ziele der Arbeit

I) Das Hauptziel der Arbeit ist die Entwicklung des Kriterienkatalogs REHBaR (**R**eporting **E**xperiments in **H**omeopathic **B**asic **R**esearch). Dieser soll als Richtlinie für Wissenschaftler dienen, die eine Arbeit aus dem Bereich Grundlagenforschung zur Homöopathie publizieren wollen. Der Katalog soll die Berichterstattung dieser Forschung standardisieren und transparent machen. REHBaR, der in einem internationalen Team von Wissenschaftlern im Konsensverfahren (Delphi-Methode) erstellt wurde, kann gleichzeitig als eine Anleitung zum wissenschaftlichen Arbeiten in der Grundlagenforschung zur Homöopathie dienen.

II) In einer voran gestellten Erhebung soll bei Wissenschaftlern, die nicht im Bereich der Grundlagenforschung zur Homöopathie forschen, die Einschätzung bestimmter Standards im Labor abgefragt werden. Dazu gehören Fragen nach Verblindung, Randomisierung, Durchführung von Kontrollen, externen und internen Replikationen sowie biometrische Betreuung. Diese Aspekte werden als Grundsätze wissenschaftlichen Arbeitens in den zu erstellenden Kriterienkatalog aufgenommen.

III) In einem letzten Teil wird überprüft, inwieweit REHBaR sich als Evaluationsinstrument für die Beurteilung publizierter Experimente aus der Grundlagenforschung zur Homöopathie eignet. Anhand des Kriterienkatalogs bewerten zwei Gutachter mehr als 30 zufällig ausgewählter Publikationen in ihrer Qualität. Die Übereinstimmungsrate wird mit zwei verschiedenen Kappa-Werten beschrieben.

## **3 Umfrage zu ausgewählten Standards in der Grundlagenforschung**

### **3.1 Der Fragebogen**

Die Umfrage hatte zum Ziel, Kernaussagen zu einer wissenschaftlichen Untersuchung im Labor zu sammeln. Es wurde ein Fragebogen mit sechs Fragen erstellt (s. Kap. 8.2), der sehr einfach und schnell zu beantworten sein sollte und an Grundlagenforscher gerichtet war, die nicht im Bereich der Homöopathie forschen. Dazu wurden insgesamt 100 Institute in Deutschland zufällig ausgewählt und angeschrieben (s. Kap. 3.2.). Der Fragebogen bezog sich auf die Themenbereiche:

1. Randomisierung: Die Versuchseinheiten werden bei jedem Experiment zufällig angeordnet.
2. Verblindung: Bei der Versuchsdurchführung weiß der Experimentator nicht, welche Behandlung (Verum oder Kontrolle/Placebo) die jeweilige Versuchsgruppe erhält. Die Substanzen werden z.B. kodiert.
3. Positiv- und Negativkontrollen: Eine so genannte Positivkontrolle wird durchgeführt, um die Reaktivität des gewählten Untersuchungsmodells zu belegen. Das Untersuchungsmodell erhält eine Gabe bzw. eine Konzentration einer Substanz, bei der in jedem Fall eine Reaktion des Systems zu erwarten ist. Die Negativkontrolle wird durchgeführt, um die Stabilität des Systems zu demonstrieren und auszuschließen, dass aufgrund der Versuchsanordnung und Durchführung Übertragungsfehler (carry-over Effekte) oder falsch-positive Ergebnisse auftreten können.
4. Interne Replikation: Das gleiche Experiment wird zu einem anderen Zeitpunkt mit neuen Materialien wiederholt.
5. Externe Replikation: Die Experimente aus einem Labor werden von anderen Personen in einem anderen Labor mit ganz neuen Materialien wiederholt.
6. Biometrische Beratung: Für Planung und Auswertung der Experimente wird ein Statistiker oder Biometriker zu Rate gezogen.

### **3.2 Auswahl der befragten Institute**

Über eine Internetrecherche wurden 100 Institute in Deutschland für die Befragung ausgewählt. Mit folgenden Stichworten wurde nach Instituten gesucht: Pflanzenphysiologie, Tierphysiologie, Grundlagenforschung, Mikrobiologie, Molekularbiologie, Pharmakologie, Toxikologie, Bi-

ochemie. Auf den Internetseiten der Institute wurde dann nach Informationen zum Forschungsschwerpunkt sowie nach Publikationslisten gesucht. Wenn hieraus ersichtlich war, dass dort keine Forschung zu Fragestellungen in der Homöopathie erfolgt, wurden der Forschungsleiter bzw. die Institutsleitung angeschrieben. Das Ablaufschema ist in der Abb. 3-1 wieder gegeben.

Die Institute wurden zum ersten Mal im September 2007 angeschrieben. Um die Rücklaufquote zu erhöhen, wurden alle im Dezember erneut angeschrieben mit der Bitte, den Fragebogen auszufüllen, falls noch nicht erfolgt. Beide Male bestand die Möglichkeit, mithilfe eines bereits frankierten Rückumschlags den Fragebogen anonym zurückzuschicken. Insgesamt gab es 46 Rückläufe. Zwei Fragebogen wurden unausgefüllt zurückschickt mit der Angabe, dass am Institut keine Grundlagenforschung zur Homöopathie durchgeführt wird; hier wurde der Fragebogen offenbar missverstanden. Ein Fragebogen wurde unbeantwortet zurückgeschickt, da der Adressat nicht mehr am Institut erreichbar war. Damit konnten 43 Antworten in die Auswertung eingeschlossen werden.

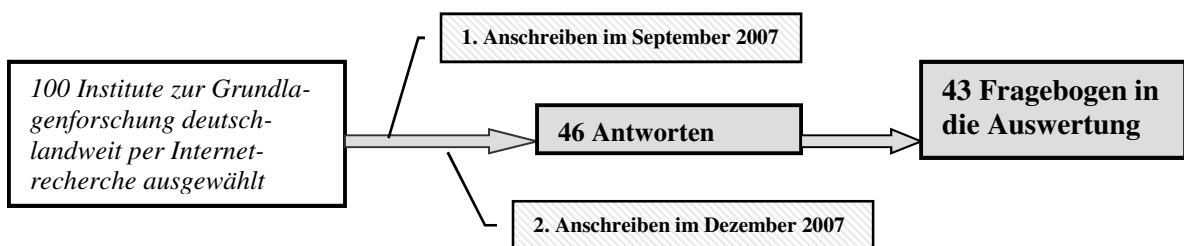


Abb. 3-1: Ablauf der Befragung schematisch dargestellt

Eine Liste aller angeschriebenen Fachbereiche und Institute findet sich im Anhang (s. Kap. 8.3). Bei der Auswahl der Institute wurde zum einen darauf geachtet, dass sich die Befragung nicht regional beschränkt, sondern Institute möglichst bundesweit erfasst wurden.

Zum anderen wurden die Institute angeschrieben, deren Internetpräsentation und Publikationslisten auf eine aktive Grundlagenforschung im jeweiligen Bereich schließen ließen. Hierbei handelte es sich am häufigsten um Institute zur Pflanzen- und zur Tierphysiologie.

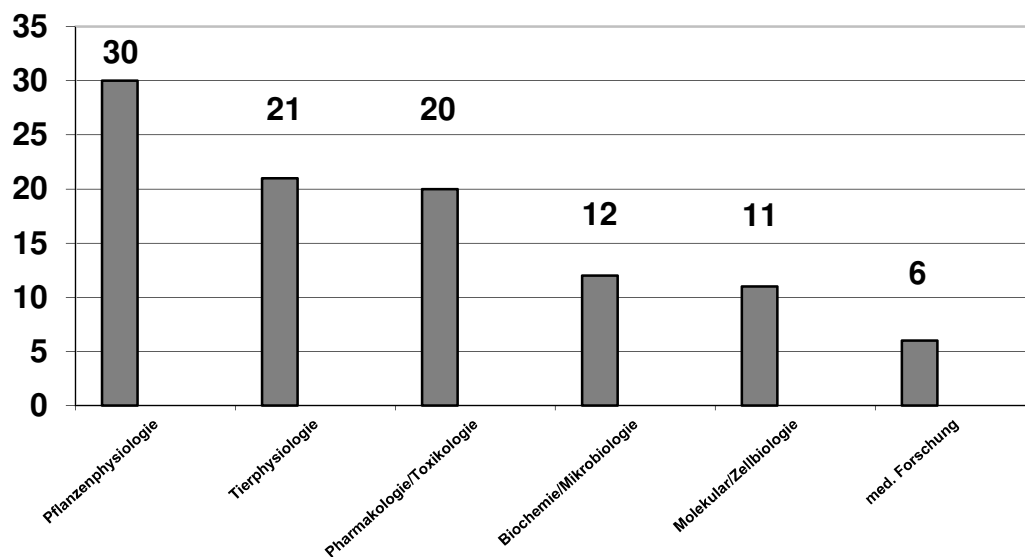


Abb. 3-2: Übersicht der Fachbereiche, aus denen Institute angeschrieben wurden (insgesamt 100 Institute).

### 3.3 Ergebnisse der Befragung

Tabelle 3-1 gibt einen Überblick zu den Ergebnissen der Fragebogenerhebung. Die Antworten werden in den nachfolgenden Kapiteln im einzelnen erläutert.

Tab. 3-1 Übersicht: Ergebnisse der Befragung (insgesamt 43 Fragebogen)

	<i>absolut notwendig</i>	<i>sehr wichtig</i>	<i>verzichtbar</i>	<i>keine Angabe</i>
Randomisierung	31	9	2	1
Verblindung	30	10	3	0
Positiv/Negativ	39	3	1	0
Interne Replikation	34	7	2	0
Externe Replikation	11	23	9	0
Statistische Beratung	12	22	8	1

#### 3.3.1 Verblindung und Randomisierung

Mehr als zwei von drei Wissenschaftlern (72 %), die geantwortet haben, randomisieren ihre Proben im Labor immer und halten dies für absolut notwendig. Nur zwei der 43 Teilnehmer

hielten das Randomisieren der Proben für verzichtbar. Die hierfür angegebenen Gründe waren bei beiden, dass aufgrund der zu untersuchenden Fragestellung und des Versuchsprotokolls eine Randomisierung nicht notwendig sei. Außerdem sei der Aufwand für eine Randomisierung zu groß, die Zahl der Versuchseinheiten oft zu gering und die zu erwartenden Effekte somit eindeutig zuzuordnen. Ähnliche Gründe fanden sich auch bei den neun Wissenschaftlern, die ihre Versuchsanordnung nur bei bestimmten Fragestellungen randomisieren.

Fast das gleiche Bild zeigt sich bei den Ergebnissen zur Frage nach der Notwendigkeit einer Verblindung (70%). Diejenigen, die nur in Ausnahmefällen verblinden, gaben als Gründe z.B. an, dass eine Verblindung bei pflanzenphysiologischen Experimenten oft nicht möglich sei oder aufgrund kleiner Probenzahlen eine „Wiedererkennung“ leicht vorkommen könne.

### 3.3.2 Positiv- und Negativkontrollen

Die Durchführung dieser Kontrollen wurde von 39 der 43 Antworten (91%) für absolut notwendig erachtet. Drei Wissenschaftler gaben an, dass sie die Positiv- und Negativkontrollen zwar für sehr wichtig erachten, sie aber nicht immer durchführen oder nur eine von beiden im Experiment ansetzen. Begründung hierfür lautete bei einem Teilnehmer, er mache sie nur, wenn er ein negatives Ergebnis erwarte.

Nur in einem Antwortfragebogen wurde „verzichtbar“ angekreuzt mit der Begründung, dass diese Art der Kontrolle in Form einer Placebogruppe bereits erbracht sei.

Einige Wissenschaftler merkten an, dass für die Wahl der geeigneten Kontrollen die Fragestellung des jeweiligen Experimentes entscheidend sei.

### 3.3.3 Replikation

Vier von fünf Wissenschaftlern (79%), die sich an der Umfrage beteiligt haben, gaben an, dass eine interne Replikation absolut notwendig sei, nur zwei hielten sie für verzichtbar. Sieben Wissenschaftler (16%) gaben an, dass sie die interne Wiederholung zwar für sehr wichtig hielten, sie aber nicht immer machten. Als Hauptgrund hierfür nannten die meisten Kosten und Machbarkeit, aber auch Tierschutzgründe wurden genannt. Darüber hinaus gaben einige Wissenschaftler an, dass die Wiederholung von Experimenten nicht nötig sei, solange die verwendete Methode valide und der Erstversuch bereits zu einem statistisch zweifelsfreien Ergebnis gekommen ist.

Die Frage nach der externen Replikation spielt bei den hier beteiligten Wissenschaftlern nur eine untergeordnete Rolle. Nur jeder vierte Wissenschaftler (11 von 43) kreuzte die Rubrik „absolut notwendig“ an. Die Hälfte der Wissenschaftler (23 von 43), die geantwortet haben, hielten



sie zwar für durchaus wichtig, aber aus Kosten- und Zeitgründen für nicht realisierbar. Ein häufig genannter Grund war hier, dass Labore nur selten über die gleiche Methodik verfügen, eine echte Wiederholung daher gar nicht machbar sei. Neun Teilnehmer (21%) gaben an, dass sie die externe Wiederholung der eigenen Experimente für überflüssig bzw. verzichtbar hielten. Als weitere Gründe wurden z.B. genannt, dass eine unabhängige Wiederholung der Experimente für eine Publikation in der Grundlagenforschung nicht nötig sei. Ein Wissenschaftler gab an, Wiederholungen nur im Rahmen von Ringversuchen durchzuführen. Ein weiterer Wissenschaftler sprach sich für externe Replikationen aus, falls die eigenen Ergebnisse strittig seien, außerdem stelle sich für ihn hier die Frage nach der Originalität der Versuche. Oft reiche es schon aus, die eigenen Daten mit bereits publizierten Daten zu vergleichen, gab ein Teilnehmer an.

### 3.3.4 Statistische Beratung

In der letzten Frage des Erhebungsbogens wurde gefragt, wie bei der Auswertung vorgegangen wird und ob eine statistische Beratung in Anspruch genommen wird. Fast 1/3 der Teilnehmer ( $n=12$ ) gab an, dass sie in jedem Fall einen Biometriker zu Rate ziehen. Ungefähr jeder zweite Wissenschaftler (51%) kontaktiert den Statistiker nur in Ausnahmefällen, acht der 43 Teilnehmer führen die Auswertungen immer selber durch.

### 3.3.5 Anmerkungen

In dieser Spalte konnten die Teilnehmer zusätzliche Angaben zu den Fragen und dem Fragebogen geben. Einer merkte an, dass in der Grundlagenforschung berücksichtigt werden müsse, ob es im jeweiligen Falle um den Nachweis eines Effektes gehe oder vielmehr um die Frage, warum der beobachtete Effekt überhaupt auftritt. Außerdem spiele das Vertrauen in das saubere Arbeiten der Mitarbeiter und anderer Wissenschaftler eine große Rolle, denn einem Protokoll müsse man vertrauen können, da man nicht alle Versuche selber durchführen könne. Ein anderer Wissenschaftler merkte an, dass man die pflanzenphysiologische Forschung nicht mit medizinischer Forschung vergleichen könne, da es bei Ersterer keine Placeboeffekte gebe.

## 3.4 Zusammenfassende Diskussion

Eine Erhebung zu ausgewählten Standards in der Grundlagenforschung außerhalb der Homöopathie wurde zum ersten Mal im Rahmen dieser Arbeit durchgeführt. Von 100 angeschriebenen Wissenschaftlern haben 43 den Fragebogen ausgefüllt. Drei zurück gesendete Fragebogen konnten nicht ausgewertet werden. Der Fragebogen war mit nur sechs Fragen kurz und sollte die generelle Einstellung von Grundlagenforschern zu Themen wie Randomisierung, Verblindung, Replikation, Kontrollen und Biometrie abfragen. Die Ergebnisse der Befragung dienten als

Grundlage für die Erarbeitung des Kriterienkatalogs REHBaR. Dabei waren aus Sicht des am Delphi-Prozess beteiligten Expertenteams folgende drei Resultate wichtig:

1. In der Grundlagenforschung wird wie in der klinischen Forschung verblindet und randomisiert. Ausnahmefälle müssen begründet sein.
2. Negativ- und Positivkontrollen gehören zu jedem Versuchsaufbau.
3. Die externe Replikation der eigenen Experimente wird von den meisten als sehr wichtig erachtet, aber kaum durchgeführt.

## 4 Erstellung des Kriterienkatalogs REHBaR

### (Reporting Experiments in Homeopathic Basic Research)

#### 4.1 Hintergrund

Im Bereich der klinischen und epidemiologischen Studien sind in den letzten fünfzehn Jahren zahlreiche Checklisten zur Verbesserung der Qualität von Publikationen entwickelt worden. Eines der ersten und am häufigsten in den Journalen empfohlenen Instrument ist CONSORT (*CONSolidated standards for Reporting Trials*), das von einem Expertenteam 1996 veröffentlicht wurde (Begg et al. 1996). Dieser Katalog wurde bislang zweimal überarbeitet (Moher et al. 2001) (Moher et al. 2005) und dient als Grundlage für Supplemente von CONSORT, die auf die Beschreibung definierter Forschungsbereiche abzielen, z.B. Studien zur Therapie mit pflanzlichen Mitteln (Gagnier et al. 2006)

RedHot (*REporting Data on HOmeopathic Treatments*) ist eine Kriterienliste, die ebenfalls auf CONSORT basiert und Kriterien enthält, die bei der Veröffentlichung von klinischen Studien zur homöopathischen Therapie zu beachten sind (Dean et al. 2007).

QUORUM ist ein Katalog, der als Checkliste für die Publikation von Metaanalysen genutzt werden kann (Moher et al. 1999). Nach der Überarbeitung im Jahr 2005 (Turpin 2005) ist die Richtlinie von einem internationalen Expertenteam weiter entwickelt und an die Erfordernisse von systematischen Reviews und Metaanalysen angepasst worden (PRISMA: *Preferred Reporting Items for Systematic reviews and Meta-Analyses*) (Moher et al. 2009).

STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*) dient als Checkliste für die Publikation von Beobachtungsstudien, Kohorten-, Fall- Kontroll- und Querschnitts- Studien (Vandenbroucke et al. 2007). Dieser Katalog wurde um Kriterien für die Berichterstattung von genetischen Assoziationsstudien erweitert: STREGA (*STrengthening the REporting of Genetic Association studies*) (Little et al. 2009).

Eine weitere Checkliste dient der Vereinheitlichung von Abstracts (Hopewell et al. 2008), von pragmatischen Studien (Zwarenstein et al. 2008) und von pharmakologischen/nicht-pharmakologischen Studien (Boutron et al. 2003).

Auch für Akupunkturstudien ist eine erprobte (Prady und Macpherson 2007) und evaluierte (Prady et al. 2008) Checkliste erarbeitet worden: STRICTA (*STandards for Reporting Interventions in Controlled Trials of Acupuncture*) (MacPherson et al. 2001).

Für eine vollständige und adäquate Beschreibung von Studien zur diagnostischen Güte wurde der Kriterienkatalog STARD (*STAndards for Reporting Diagnostic accuracy studies*) entwickelt (Bossuyt et al. 2003).

Mit Hilfe von SQUIRE (*Standards for Quality Improvement Reporting Excellence*) soll die Qualität der Berichterstattung von Studien im Bereich des Gesundheitswesens verbessert werden (Davidoff et al. 2009).

Eine Zusammenstellung aller existierenden Richtlinien für die Erstellung von Publikationen von klinischen Studien findet sich bei EQUATOR (*Enhancing the Quality and Transparency Of health Research*). Diese neue Initiative bietet eine Übersicht und damit die Möglichkeit, die geeignete Richtlinie für den jeweiligen Autor zu finden (Simera et al. 2009). Alle diese Richtlinien sind als Empfehlungen zu verstehen, die z.B. von wissenschaftlichen Journalen den Autoren vorgegeben werden.

In den Biowissenschaften hat sich eine Initiative gebildet, bei der an Standards für die Wiedergabe von Daten aus Experimenten gearbeitet wird. MIBBI (*Minimum Information for Biological and Biomedical Investigations*) ist die Dachorganisation dieser Arbeitsgruppen, die für verschiedene, spezielle Fragestellungen Vorschläge entwickelt haben, z.B. wie eine große Datenmenge optimal publiziert wird (Taylor et al. 2008). Unter [www.mibbi.org](http://www.mibbi.org) finden sich alle Unterorganisationen mit den jeweiligen Richtlinien.

Das Ziel der vorliegenden Arbeit ist die Entwicklung eines Kriterienkatalogs, der als Richtlinie Autoren bei der Berichterstattung von Experimenten aus der Grundlagenforschung zur Homöopathie dienen soll.

## **4.2 Der Delphi-Prozess als Konsensumethode**

Für die Erstellung dieses Kriterienkatalogs im Rahmen einer Abstimmungsprozesses wurde die Delphi-Methode gewählt. Die Methodik eines Delphi-Prozesses sieht vor, dass ein Expertenteam in einem strukturierten Beurteilungs- und Abstimmungsverfahren die wichtigsten Informationen und Entscheidungskriterien zusammenträgt, die für die Beantwortung einer Frage oder einer gemeinsamen Vorgehensweise erforderlich sind (Linstone und Turoff 2002). In Form von Fragebögen haben die ausgewählten Teilnehmer in diesem Prozess die Möglichkeit, zum einen vorgegebene Parameter zu beurteilen, zum anderen eigene Empfehlungen und neue Aspekte mit einzubringen.

In der Medizin wurde z.B. eine Behandlungsrichtlinie für die Therapie der Endometriose mit Phytotherapeutika aus der chinesischen Medizin von einem Expertenteam mithilfe dieser Konsensumethode erarbeitet (Flower et al. 2007). Verhagen et al. haben einen Kriterienkatalog in

einem Team von Experten entwickelt, der zur Beurteilung der Qualität von randomisierten klinischen Studien im Rahmen von systematischen Reviews dienen soll (Verhagen et al. 1998).

In der vorliegenden Arbeit wurde der Delphi-Prozess in insgesamt fünf Runden, davon zwei Konsensuskonferenzen durchgeführt. Die Auswahl der Teilnehmer erfolgte unter zwei Gesichtspunkten: Erstens sollten es Wissenschaftler sein, die in den letzten fünf Jahren Experimente zur Homöopathie publiziert hatten. Zum zweiten sollten aus Gründen der Praktikabilität nur Wissenschaftler aus dem europäischen Raum eingebunden werden. Es wurden folgende Wissenschaftler zu der Delphi-Runde eingeladen:

Die in Europa zur Zeit führenden Grundlagenforscher in der Homöopathie **Dr. S. Baumgartner** von der Kikom Universität Bern/Schweiz, **Dr. Lucietta Betti** aus Bologna/Italien, **Dr. Christian Endler** aus Graz/Österreich sowie **Dr. Roeland van Wijk** aus Utrecht/Niederlande. Dr. Baumgartner und Dr. Betti arbeiten mit Pflanzen und haben in den letzten Jahren über Wirkungen homöopathischer Substanzen auf das Pflanzenwachstum geforscht und publiziert. Dr. Endler beschäftigt sich mit der Evolutionsbiologie insbesondere bei Amphibien, während Dr. van Wijk im Bereich biophysikalischer Grundlagenforschung zum homöopathischen Simile-Prinzip arbeitet.

Darüber hinaus konnten zwei Wissenschaftler für die Expertenrunde gewonnen werden, die sowohl auf dem Gebiet der Grundlagenforschung zur Homöopathie als auch in der Bewertung klinischer Studien geforscht haben: **PD Dr. Klaus Linde**/ TU München und **Prof. Dr. Claudia Witt**/ Charité Berlin. Klaus Linde hatte den ersten Bewertungskatalog im Jahr 1994 zu den toxiologischen Experimente entwickelt. In den letzten Jahren ist er zum Experten von Metaanalysen und systematischen Reviews in der klinischen Forschung geworden (Linde et al. 1997). Claudia Witt hat in zwei Arbeiten den SAPEH Score entwickelt und mit diesem einen Review über biophysikalische (Becker-Witt et al. 2003) und in-vitro- Experimente (Witt et al. 2007) publiziert.

Für den Delphi-Prozess stellte sich außerdem zur Verfügung: **Dr. Henning Albrecht**, der gemeinsam mit Dr. Roeland van Wijk eine Datenbank erstellt hat, in der alle publizierten Experimente, die zugänglich sind, aus dem Bereich der Grundlagenforschung zur Homöopathie verwaltet werden (van Wijk und Albrecht 2007). **Dipl.-Stat. Rainer Lüdtke** konnte für das Expertenteam gewonnen werden, um die biometrischen Aspekte in den Katalog mit einzubringen. Hierbei ging es vor allem um die Anforderungen an die Auswertungskonzepte und statistische Verfahren in einer Publikation. Darüber hinaus war er bei der Entwicklung und Publikation der beiden Reviews von Witt et al. beteiligt. **PD Dr. Frauke Musial** (Universität Duisburg-Essen)

## Erstellung des Kriterienkatalogs REHBaR (Reporting Experiments in Homeopathic Basic Research)

---

hat sich auf dem Gebiet der Grundlagenforschung zur Biopsychologie habilitiert und sollte das Team mit der Sicht auf generelle Fragen in der Grundlagenforschung ergänzen.

In der **Vorbereitungsphase** zum Delphi-Prozess wurden 58 Kriterien in Form eines Fragenkatalogs zusammengefasst (s. Kap. 8.4). Grundlage hierfür bildeten die Kriterien in den bereits publizierten Scores aus Linde et al. (1994) und Becker-Witt et al. (2003). Die Liste wurde noch um weitere Kriterien ergänzt, die sich aus Vorabgesprächen mit Experten auf dem Gebiet als notwendig herausgestellt hatten. Ziel dieser ersten Auflistung war es, möglichst ausführlich alle Aspekte zu erfassen.

In der ersten Runde (**Delphi I**) sollten die Experten eine Bewertung der vorgeschlagenen Kriterien übernehmen. Ziel war es, die Zahl der Kriterien auf diejenigen zu reduzieren, die von allen als absolut notwendig erachtet werden. Die Teilnehmer hatten auch die Möglichkeit, fehlende Kriterien zu ergänzen. Dazu sollten die Teilnehmer mit Hilfe einer fünfstufigen Skala – der „5-Point-Likert-Scale“ – (Verhagen et al. 1998) jede der im Fragenkatalog formulierte Frage wie folgt bewerten:

Absolutely necessary	(4 points)
Moderately necessary	(3 points)
Neutral	(2 points)
Moderately unnecessary or disagree	(1 point)
Strongly unnecessary or disagree	(0 point)

Dazu gab es noch die Möglichkeit eventuell fehlende Aspekte zu ergänzen. Die abgegebenen Einschätzungen wurden wie folgt ausgewertet: Pro Frage wurden die abgegebenen Punkte addiert und durch die Zahl der Teilnehmer (neun) dividiert. Danach ergab sich folgende Klassifikation der 58 Aspekte:

18 Fragen	Absolutely necessary (3,5 - 4)
31 Fragen	Moderately necessary (2,5 - 3)
9 Fragen	Neutral (2)

Nur eine Frage bzw. ein Aspekt wurde ergänzt und für die Diskussion in der Konferenz mit aufgenommen. Die Fragen des Katalogs wurden als Diskussionsgrundlage in die drei Kategorien (s. oben) eingeordnet mit dem Ziel, vor allem die Punkte, die mit „neutral“ oder „moderately necessary“ bewertet wurden, in den Katalog abschließend aufzunehmen oder zu verwerfen.

Die erste Konsensuskonferenz (**Delphi II**) fand im Oktober 2007 statt. Hier sollten die ausgewählten Beurteilungskriterien diskutiert werden in Hinblick auf die Erstellung eines Bewer-

tungsscores, vergleichbar mit dem SAPEH-Score (Witt et al. 2007) oder dem Jadad-Score (Moher et al. 1995). Nach der Diskussion der Kriterien waren sich die Teilnehmer einig, dass im ersten Schritt eine Kriterienliste als Standard für die Erstellung einer Publikation erarbeitet werden sollte. Erst im zweiten Schritt könne ein Instrument für die Bewertung einer Publikation konzipiert werden. Die Diskussion, welches Instrument in der Grundlagenforschung zur Homöopathie für die Beurteilung der Qualität geeignet sei, sollte erst bei der folgenden Konferenz diskutiert werden.

Da für die Publikation klinischer Studien bereits einige Kriterienkataloge entwickelt wurden (s. oben), sollten diese als Vorlage für den zu erstellenden Katalog für die Grundlagenforschung zur Homöopathie dienen: CONSORT (Moher et al. 2001) und STROBE (Vandenbroucke et al. 2007). Der Kriterienkatalog sollte international publiziert werden und wurde somit auf Englisch formuliert.

Der Katalog wurde mit den beschlossenen Kriterien in **Delphi III** erarbeitet und zur Vorbereitung auf die nächste Konsensuskonferenz an alle Teilnehmer verschickt. Neben der Möglichkeit, die Formulierungen und Kriterien zu überarbeiten, sollte überprüft werden, welche Angaben im Katalog fehlen, um ihn sowohl für die Beschreibung konfirmatorischer als auch explorativer Experimente nutzen zu können.

Im Dezember 2007 fand die zweite Konsensuskonferenz **Delphi IV** statt mit dem Ziel, den überarbeiteten Katalog abschließend zu diskutieren. Darüber hinaus sollte beraten werden, ob der Katalog in der vorliegenden Form für die Beurteilung der Güte einer Publikation ausreicht oder ob ein Bewertungsscore besser geeignet wäre. Während der Diskussion waren sich die Teilnehmer schnell darin einig, dass ein skaliertes Score nicht der Qualitätsbeurteilung gerecht werden kann. Der Kriterienkatalog zur Erstellung einer Publikation sollte mit weiteren Unterteilungen und Spalten zum Ankreuzen adaptiert werden, um ihn als Beurteilungsinstrument nutzen zu können. Dazu sollte ein ausführlicher Erläuterungskatalog für jedes Kriterium geschrieben und geeignete Beispiele für die einzelnen Vorgaben aus der Literatur heraus gesucht werden.

In der letzten Delphi-Runde (**Delphi V**) wurden sämtliche Kriterien für die erste Tabelle noch einmal überarbeitet. Darüber hinaus wurde zu jedem Kriterium eine Erläuterung geschrieben sowie ein passendes Beispiel aus vorhandenen Publikationen herausgesucht. Dieses Dokument wurde zur Abstimmung an alle Teilnehmer verschickt und im Anschluss daran bei einem wissenschaftlichen Journal zur Publikation eingereicht (Stock-Schroer et al. 2009).

Ein Überblick über den gesamten Ablauf des Prozesses findet sich in Abb. 4-1. Die Methode des Delphi-Prozesses ist in einer weiteren Publikation veröffentlicht worden (Stock-Schroer et al. 2009).

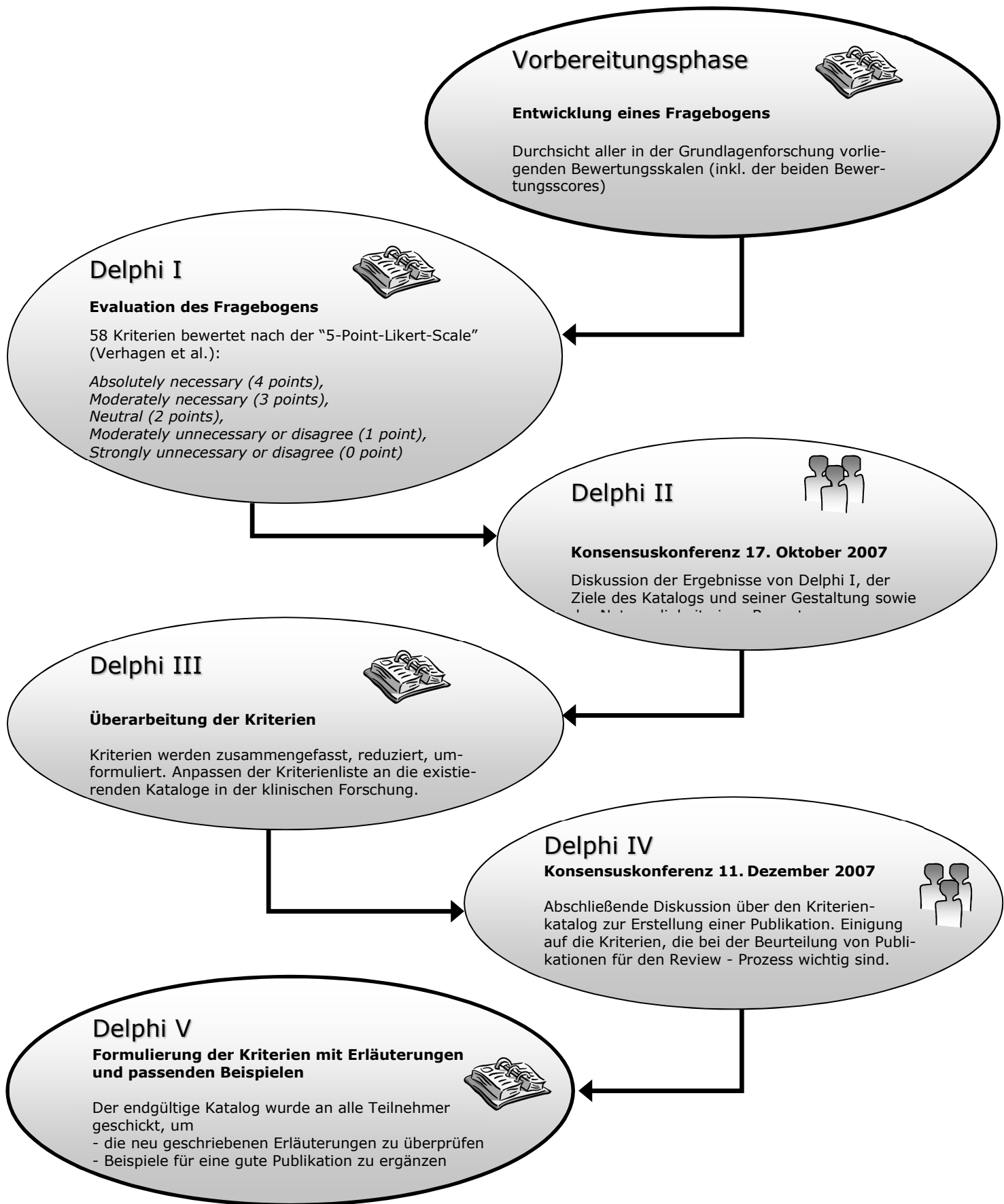


Abb. 4-1: Ablaufschema des Delphi-Prozesses (Oktober 2007 bis März 2009)



### 4.3 Der Kriterienkatalog

**Tab. 4-1: Kriterien, die bei der Berichterstattung von Experimenten aus der Grundlagenforschung zur Homöopathie erfüllt sein müssen: Reporting Experiments in Homeopathic Basic Research (REHBaR) aus (Stock-Schroer et al. 2009)**

Item	N <sup>o</sup>	Description
Title	1	Title indicates the experimental model and the intervention
Abstract	2	Abstract provides an informative and balanced summary of what was done and found
<b>Introduction</b>		
Background	3	Scientific background, presentation of experimental model(s), explanation of rationale, including homeopathic principles (e.g. similia principle, potentisation, proving) and type of homeopathy (isopathy, classical vs complex homeopathy)
Objectives/Hypotheses	4	Objectives and hypotheses with outcome measures. For confirmatory experiments: specific hypotheses and clearly defined primary outcome measure. For exploratory experiments: hypotheses inducing the investigations.
<b>Materials and Methods</b>		
Materials	5	Detailed description of all used materials (e.g. biological system, devices, substances, instruments)
Materials (homeopathy specific)	6	Manufacturer, Pharmacopoeia (or process) of medications, potency and steps of dilution, dilution method, substance starting point of dilution (e.g. mother tincture, D1, nosode)
Homeopathic Controls	7	Precise details on the preparation of the control substance
System Performance Controls	8	Report on negative and positive controls
Quality control	9	Procedures and efforts used to enhance the quality and reliability of the experimental procedure
Object of investigation	10	Selection criteria for the particular system used: in vivo, in vitro, biological, physical, biochemical
Experimental Setup	11	Detailed description of experimental conditions and procedure
Replication	12	If experiment has internal replications, detailed description is given of which materials were reused and which have been changed
Parameters	13	All measured parameters described in detail
Intervention	14	Precise details of the interventions intended for each group and how and when they were actually administered
Allocation	15	Method used to generate the group allocation including details (e.g. randomisation, blocking, stratification)
Blinding	16	Description if any procedures or interventions were concealed (if yes, details given)
Statistical Methods	17	Statistical tests and procedure of calculation are described: Methods for additional analyses like adjusted analyses
<b>Results</b>		
Numbers analysed	18	Number of experiments with exact number of treated units per setting which were included in each analysis and reporting missing samples, drop outs
Data (descriptive)	19	Results are given in tables or figures showing mean or median together with variability (e.g. SD and /or range) for absolute data (and differences)
Data (inferential)	20	Gives appropriate measures of effect size, uncertainty and probability
<b>Discussion</b>		
Interpretation	21	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision
Evidence	22	General interpretation of results in the context of current evidence. Discuss the generalisability /external validity of the study results
Experimental Model	23	Explanation why this model, these parameters were chosen and its adequacy for answering the questions including homeopathic aspects

In Tabelle 4-1 ist der Kriterienkatalog mit den dazugehörigen kurzen Erläuterungen als Ergebnis des Delphi-Prozesses dargestellt. Er dient als Anleitung beim Schreiben wissenschaftlicher Publikationen und soll dem Autor helfen, alle relevanten Informationen wieder zu geben.

Zu jedem einzelnen Kriterium aus dem Katalog wurde eine ausführliche Erläuterung geschrieben und ein passendes Beispiel aus einer vorhandenen Publikation hinzugefügt (Stock-Schroer et al. 2009). In Kap. 8.5 findet sich der publizierte Katalog auf Englisch, im Nachfolgenden werden die einzelnen Kriterien auf Deutsch erläutert.

### **(1) Titel**

Der Titel einer Publikation sollte so präzise formuliert sein, dass aus ihm ersichtlich wird, welche Untersuchungen in der Publikation erörtert werden. Der Titel muss informieren und das Interesse eines Lesers wecken, ob die Publikation in sein Arbeits- oder Interessensgebiet fällt. Dazu sollte der Titel das verwendete Untersuchungsmodell genau benennen und angeben, ob die Experimente “in vivo” oder “in vitro” durchgeführt wurden. Darüber hinaus sollte zu erkennen sein, wie und mit welchen Substanzen behandelt wurde.

### **(2) Zusammenfassung/Abstract**

Das Abstract bietet ausführlichere Informationen, aus denen der Leser auf die Bedeutung der Publikation für das eigene Arbeitsgebiet schließen kann. Es gibt einen Überblick über Fragestellung, Methodik, Ergebnisse und Diskussion der publizierten Arbeit und sollte alle wesentlichen Merkmale der Arbeit enthalten. Es wird empfohlen, das Abstract in Überschriften nach den oben genannten Stichworten zu gliedern. Folgende Angaben müssen enthalten sein: Die exakte Bezeichnung der verwendeten Substanzen, der Behandlung (z.B. Strahlung, Temperatur), die Art der Applikation (z.B. oral oder Injektion), die Dosis (z.B. ml,  $\mu\text{mol}$ ) sowie die verwendeten Konzentrationen bzw. Potenzen der homöopathischen Arzneimittel (D- oder C- Potenzen usw.). Im Abstract sollten die wichtigsten Ergebnisse zusammengefasst und diskutiert werden unter Angabe der zugrunde liegenden Zahlen sowie Schätzung für korrelative Zusammenhänge, Variabilität und Wahrscheinlichkeiten.

### **(3–4) Einleitung**

#### **(3) Hintergrund**

Die Einleitung gibt einen Überblick zu dem Forschungsgebiet, in das die publizierte Arbeit fällt. Hier sollten alle für die untersuchte Fragestellung relevanten Forschungsarbeiten zitiert werden, inklusive der eigenen Arbeiten auf dem Gebiet. Aus der Einleitung muss sich dem Leser die Notwendigkeit der aktuellen Fragestellung erschließen. Ebenso finden sich hier die Begründung

für das gewählte Untersuchungsmodell, für den Hauptzielparameter und warum gerade diese beiden für die Beantwortung der Fragestellung geeignet scheinen. Die Homöopathie betreffend ist es wichtig anzugeben, welches homöopathische Wirkprinzip den Experimenten zugrunde liegt oder vermutet wird: Simile-Prinzip, Wirkung des Potenzziervorgangs oder Arzneimittelpri-  
fung. Ebenso muss gekennzeichnet werden, um welchen Bereich der Homöopathie es sich bei dieser Publikation handelt: Isopathie, klassische Homöopathie (nach den Grundprinzipien der Homöopathie verabreichte Einzelmittel) oder Komplexmittel-Homöopathie (mehr als zwei ho-  
möopathische Arzneimittel werden gemeinsam gegeben).

#### **(4) Ziel und Hypothesen**

Neben der Einführung in das Themengebiet sollte am Ende der Einleitung die zugrunde liegen-  
de Hypothese oder das Ziel der publizierten Arbeit formuliert sein. Dabei ist zu beachten, um  
welche Art von Experimenten es sich in der Publikation handelt. Bei einem konfirmatorischen  
Experiment müssen eine eindeutige Hypothese sowie klar definierte Zielparameter beschrieben  
sein. Bei explorativen Experimenten ist solch eine Definition oft nicht möglich. Hier genügt ei-  
ne Erläuterung, welche Vorüberlegungen und Ergebnisse von Vorstudien für das Experiment  
relevant waren.

#### **(5–17) Material und Methoden**

Dieser Teil der Publikation entspricht einer Anleitung, um eine möglichst exakte Wiederholung  
(Replikation) der Versuche zu ermöglichen. Deshalb sollten Material und Methoden so detail-  
liert wie möglich beschrieben werden. Falls in dem Journal nicht genügend Raum für ausführli-  
che Beschreibungen vorgesehen ist, wird empfohlen, mittels eines Anhangs oder Links auf eine  
eigene Homepage diese zusätzliche Informationen zur Verfügung zu stellen.

#### **(5) Materialien**

Eine detaillierte Beschreibung des im Experiment untersuchten biologischen, physikalischen  
oder biochemischen Untersuchungsmodells ist notwendig, wenn möglich und vorhanden mit  
Nennung des exakten Namens (z.B. Wachstumsmodell, Intoxikations-Modell). Sämtliche Gerä-  
te, Verbrauchsmaterialien und Messapparaturen sollten beschrieben sein, mit genauer Angabe  
aller technischer Daten inklusive Firmenname, Seriennummer usw.. Ist ein eigenes Gerät, Un-  
tersuchungsmodell oder ähnliches entwickelt oder verwendet worden, muss es hier spezifiziert  
werden. Eine exakte Angabe zur Anzahl der untersuchten Proben (z.B. Zellen, Organismen)  
sowie die absolute Zahl der Experimente sollte der Übersicht halber möglichst in einem Ablauf-  
diagramm wiedergegeben werden.

### **(6) Material (Homöopathie-spezifisch)**

Die Herstellung bzw. der Bezug der homöopathischen Materialien sollte detailliert beschrieben bzw. benannt werden. Wenn bekannte Herstellungsmethoden verwendet werden, dann sollten diese genannt werden, z.B. nach dem homöopathischen Arzneibuch (HAB) oder nach der Korsakov-(Einglas)-Methode. Wenn die homöopathischen Präparationen nach einer eigenen Methode hergestellt wurden, muss diese Abweichung begründet werden. Alle Schritte sollten im Einzelnen genau beschrieben sein, so z.B. die Zahl der Schläge bei der Verschüttelung, die Frequenz und Richtung dieser (horizontal, vertikal), ob gegen eine Unterlage oder in der Luft, maschinell oder per Hand usw.

Es wird empfohlen, den Ausgangspunkt der homöopathischen Präparationen und die exakte Bezeichnung der verwendeten Substanz: z.B. Urtinktur, eine D1 oder Nosode (homöopathisch aufbereitete Mittel aus pathologischem Material) zu nennen. Wenn die Substanzen kommerziell bezogen wurden, sollten Name und Bezugsadresse gegeben werden. Falls die verwendeten Substanzen gelagert werden mussten, sind Angaben erforderlich, wo und wie gelagert wurde (im Dunkeln, in der Kühlung) und für welchen Zeitraum.

### **(7) Kontrollen (Homöopathie-spezifisch)**

Die Wahl und Herstellungsweise der Kontrolle spielt in der Homöopathie eine bedeutende Rolle und muss genau beschrieben sein. Dazu gehört die Begründung, warum die jeweilige Kontrolle für die Fragestellung verwendet wurde.

Beispiele für adäquate Kontrollen sind:

1. Wenn der zu erwartende Effekt auf den Verschüttelungsprozess an sich zurück zu führen ist, ist eine unpotenzierte Kontrolle zu wählen.
2. Wenn ein Effekt von der potenzierten Substanz selbst erwartet wird, ist eine potenzierte Lösung ohne zugegebene Substanz oder eine in der gleichen Potenzstufe hergestellte andere Substanz eine adäquate Kontrolle.
3. Wenn das Experiment die einzelnen Potenzierungsschritte und ihre Effekte zum Untersuchungsgegenstand hat, dann ist eine Kombination von potenziertem Lösungsmittel und einer verdünnten Urtinktur die richtige Wahl.
4. Wenn im Experiment das "Simile-Prinzip" untersucht werden soll, sind die Substanzen als Kontrollen zu wählen, die nach diesem Prinzip in unterschiedlichem Ausmaß Ähnlichkeit mit dem Krankheitsbild aufweisen.

### **(8) Positiv- und Negativ-Kontrollen**

Positiv- und Negativ-Kontrollen sind in der Grundlagenforschung die Standardkontrollen. Welche der Kontrollen bei Experimenten zu homöopathischen Fragestellungen zu den Positiv- bzw. Negativ-Kontrollen gerechnet werden, sollte beschrieben sein. Wurden sie nicht durchgeführt, muss dieses begründet werden.

1. Positiv-Kontrollen durchlaufen den gleichen Versuchsablauf, um die Reaktivität des Systems zu dokumentieren. Wie viele Proben mit welcher Substanz oder unter welchen Bedingungen zu welchem Ergebnis gekommen sind, sollte angegeben werden.
2. Negativ-Kontrollen dienen dazu "carry-over" Effekte auszuschließen (zeitverzögerte Reaktionen, die nach der Behandlung auftreten können). In einigen Versuchssystemen sind sie notwendig, um die Stabilität des „unbehandelten“ Systems zu prüfen (inaktiviertes Stadium z.B. von menschlichen Basophilen).
3. Eine Reihe an Negativ-Kontrollen (sogenannte blind-runs; systematic negative control experiments) sollten systematisch durchgeführt werden, um neben Störungen im Versuchsablauf falsch-positive Ergebnisse auszuschließen. Artefakte, die durch Unterschiede in der Versuchsanordnung evtl. auftreten, können zu solchen falsch-positiven Ergebnissen führen: minimale Temperatur- und Lichtschwankungen oder zeitlich bedingte Unregelmäßigkeiten, z.B. wenn einige Einheiten wesentlich früher oder später behandelt bzw. gemessen werden.

### **(9) Qualitätskontrolle**

Der Versuchsleiter muss für die wissenschaftliche Qualität des Experimentes garantieren. Dazu sollten Angaben in der Publikation gemacht werden, welche Schritte zur Qualitätssicherung unternommen wurden: Ausbildung und Qualifikation der Mitarbeiter, Kalibrierung und Wartung der Geräte, Supervision der Mitarbeiter, Maßnahmen, um Verunreinigungen bei Substanzen und dem zu prüfenden System zu verhindern, usw.

### **(10) Untersuchungsmodell**

Eine detaillierte Beschreibung des gewählten Untersuchungsmodells ist notwendig. Dazu ist die genaue Bezeichnung erforderlich: Tier, Gattung, Spezies, Geschlecht, Organ, Zellkompartiment (Mitochondrien, Nukleus), Name der verwendeten Zelllinie, Herkunft der verwendeten Zellen, lateinischer Pflanzename, verwendete Teile der Pflanze, Pilze, Bakterien, Virus, physikalisches System (Spektroskopie), biochemisches Modell (Enzyme), chemisches Modell, usw. Neben der Beschreibung ist auch eine Begründung zu geben, warum das gewählte Untersuchungsmodell für die Beantwortung der Fragestellung geeignet erscheint.

### **(11) Versuchsaufbau**

In der Pflanzenforschung ist es wichtig, Angaben zu allen relevanten Wachstumsbedingungen zu geben: Erde, Substrat, Zeitraum des Wachstums, Jahreszeit der Versuchsdurchführung, Zeitpunkt der Ernte, Herkunft, Lagerung der Samen, Früchte usw.. Bei Tierexperimenten dürfen Angaben zu Geschlecht, Alter, Gewicht, Fütterung und allgemeine Haltungsbedingungen nicht fehlen. Wenn in-vitro-Experimente (z.B. mit Enzymen, physikalische Parameter) durchgeführt wurden, sollte jeder Faktor, der möglicherweise einen Einfluss auf das Experiment genommen hat, beschrieben sein. Dazu gehören auch Angaben über die Herstellungsverfahren, Herkunft der Proben, Inkubationszeit und Reinigungs- oder Aufbereitungstechniken. Jede Maßnahme zur Standardisierung der physikalischen Bedingungen muss beschrieben werden (z.B. Raumfeuchte, Temperatur, Licht).

Ein exakter Plan des Versuchsaufbaus mit einer schrittweisen Beschreibung des Versuchsablaufs z.B. in Form eines Ablaufdiagramms wird empfohlen. Die exakte Bezifferung der Proben pro Experiment und die genaue Zahl der durchgeführten Experimente muss an dieser Stelle der Publikation angegeben werden. Falls der Versuchsaufbau oder -ablauf in einer früheren Publikation ausführlich beschrieben wurde, kann mit Hinweis auf die Quelle in der vorliegenden Publikation darauf verzichtet werden.

### **(12) Replikation**

Wenn die Experimente intern und unabhängig voneinander wiederholt wurden, ist die Zahl der Wiederholungsversuche zu nennen. Darüber hinaus ist es wichtig anzugeben, welche der Substanzen und Materialien wieder verwendet bzw. erneuert wurden. Falls es sich um eine externe Replikation handelt, sind die Abweichungen vom vorangegangenen Versuch zu beschreiben.

### **(13) Messparameter**

Die Reaktion des Untersuchungsmodells wird anhand eines oder mehrerer messbarer Zielparameter bestimmt. Welche Messgrößen in welchen Einheiten gemessen wurden, ist klar zu definieren. Darüber hinaus ist eine Begründung zu geben, warum der jeweilige Parameter für die Beantwortung der Fragestellung geeignet erscheint.

### **(14) Behandlung/Applikation**

Die Beschreibung der Behandlungen sollte folgende Fragen beantworten können:

1. Dosis: Welche Menge wurde in welcher Konzentration verabreicht?
2. Zeitpunkt und Häufigkeit: Wann wurde die Testsubstanz (Kontrollsubstanz) zum ersten Mal verabreicht, wie oft und in welchen Zeitabständen und zu welchen Tageszeiten?

3. Darreichungsform: In welcher Konsistenz erfolgte die Verabreichung? Globuli, Tabletten, Flüssigkeit?
4. Applikationsmethode: Wie wurden Kontroll- und Testsubstanz verabreicht? Oral? Per Injektion?

### **(15) Randomisierung**

Die Proben sollten in zufälliger Anordnung mit der Test - oder Kontrollsubstanz bzw. unter den zu vergleichenden Bedingungen behandelt werden. Hierzu ist es notwendig, exakt zu beschreiben, wie die Zufallsverteilung in die jeweilige Gruppe erfolgte. Wenn die Zuteilung randomisiert erfolgte (Einzel - oder Blockrandomisation), dürfen die Angaben, wie (per Los, Verwendung einer speziellen Software) und von wem die Randomisierungsliste erstellt wurde, nicht fehlen. Ist keine Randomisierung erfolgt, muss dies explizit genannt und begründet werden.

### **(16) Verblindung**

Die Art und Weise wie der Experimentator verblindet wurde, d.h. bei der Versuchsdurchführung nicht wusste, welche Gruppe welche Behandlung erhielt, ist detailliert zu beschreiben. Auch wie die Entblindung hinterher erfolgte und wo die Randomisationsliste verwahrt wurde, ist zu benennen. Ist eine Verblindung nicht möglich gewesen oder nicht durchgeführt worden, muss dies begründet werden.

### **(17) Statistik**

Eine genaue Darstellung der verwendeten statistischen Verfahren und Tests sind in der Publikation anzugeben, ebenso eine Erläuterung, warum diese Verfahren für geeignet befunden wurden. Dazu gehören Angaben, in welchen Fällen eine Adjustierung für multiples Testen vorgenommen wurde, wie Confounder (intervenierende Variablen, die systematisch den Effekt beeinflussen) und Störfaktoren, die sowohl systematisch als auch unsystematisch sein können, und wie fehlende Werte (drop-outs) in der Analyse berücksichtigt wurden.

### **(18–20) Ergebnisse**

#### **(18) Auswertung**

Jeder Auswertungsschritt sollte im Einzelnen dokumentiert werden. Dazu gehören genaue Angaben, welche und wie viele Werte in die Rechnung mit eingeschlossen sind, und eine Begründung, warum bestimmte Ergebnisse von der Auswertung ausgeschlossen wurden. Ein- und Ausschlussgründe sollten dargestellt und begründet werden, z.B. wenn Ergebnisse ausgeschlossen werden, die auf eindeutige Bedienungs - oder Behandlungsfehler zurückzuführen sind. Fehlende

Werte, die auf einen so genannten „drop-out“ zurück zu führen sind, müssen in der statistischen Auswertung mit berücksichtigt oder diskutiert werden.

### **(19) Deskriptive Datenanalyse**

Neben einer kurzen beschreibenden Darstellung der Ergebnisse sollten diese detailliert und übersichtlich in Tabellen zusammengefasst werden. Abbildungen sind geeignet, um dem Leser die zentralen Ergebnisse zu veranschaulichen. Bei kontinuierlichen Variablen empfehlen sich zwei Darstellungsweisen: Bei einer Normalverteilung der Daten wird der Mittelwert zusammen mit der Standardabweichung dargestellt. Bei nicht-normal verteilten Daten wird der Median, die erste und dritte Quartile sowie der größte und der kleinste Wert (Range) aufgetragen. Bei binären oder kategorialen Variablen sollten zur besseren Übersicht neben absoluten Häufigkeiten auch Prozentwerte dargestellt werden.

### **(20) Inferenzielle Datenanalyse**

Die Effektgröße, d.h. die mittlere Differenz der Ergebnisse von Test- und Kontrollgruppe, sollte gemeinsam mit dem Konfidenzintervall oder dem Standardfehler des jeweiligen Effektes angegeben werden. Daneben ist der p-Wert als Maß für die Wahrscheinlichkeit, dass tatsächlich ein Unterschied vorliegt, anzugeben. Neben der Berechnung der Effektgrößen sollten alle ermittelten p-Werte angegeben werden, auch wenn sie nicht signifikant sind. Angaben zur Signifikanz und Wahrscheinlichkeit sollten auch im Text erläutert werden und nicht nur in Tabellen und Abbildungen gelistet sein. Bei konfirmatorischen Experimenten sind Informationen zur Kalkulation der Power und somit der Fallzahlschätzung notwendig.

### **(21–23) Diskussion**

#### **(21) Interpretation**

Die Interpretation sollte sich auf die im Ergebnisteil dargestellten Resultate beziehen. Dabei sind die in der Einleitung beschriebenen Vorüberlegungen (Hypothesen, Erwartungen) zu der Studie aufzugreifen und mit den tatsächlich gefundenen Daten und Effekten zu vergleichen und zu diskutieren. Alle einschränkenden Faktoren der Experimente wie eine mögliche Verzerrung der Ergebnisse (Bias), Fehler bei der Versuchsdurchführung u.ä. sollten ebenso wie statistische Mängel diskutiert werden.



## **(22) Evidenz**

Nach Interpretation und Diskussion der eigenen Resultate müssen diese in Beziehung zu bereits publizierten Ergebnissen gesetzt werden. Dabei sollte diskutiert werden, ob sich aus allen vorliegenden Erkenntnissen eine valide Aussage zur Fragestellung ableiten lässt.

## **(23) Untersuchungsmodell**

Eine kritische Einschätzung des verwendeten Untersuchungsmodells wird empfohlen. Dies beinhaltet eine Erläuterung, ob das gewählte Untersuchungsmodell und die Messparameter zur Beantwortung der Fragestellung/Hypothese geeignet waren. Bezüglich des Forschungsgebietes Homöopathie ist an dieser Stelle auch eine Diskussion über die Aussagekraft bezüglich des untersuchten homöopathischen Prinzips erforderlich, z.B. ob der gefundene Effekt auf das „Simile-Prinzip“ zurück zu führen ist oder eine isopathische Wirkung erklärt.

## **4.4 Diskussion und Bedeutung des Kriterienkatalogs**

Gemeinsam mit einem Team von neun Wissenschaftlern aus dem Bereich der klinischen Forschung und der Grundlagenforschung zur Homöopathie wurde in einem über 18 Monate andauernden Delphi-Prozess ein umfangreicher Katalog an Kriterien erarbeitet.

Die Delphi-Methode ist für die Erstellung eines Kriterienkatalogs geeignet. Mit neun Experten war die Gruppe hinreichend groß und der Delphi-Prozess mit mehreren Abstimmungsrunden ausführlich genug, um einen gut abgestimmten Konsens zu erzielen. Das Übereinkommen innerhalb eines Expertenteams hängt ganz entscheidend von den Teilnehmern, von dem Vorgehen bei der Zusammenfassung der Ergebnisse, deren Präsentation und Einigung bei möglichen Konflikten ab (Kottner et al. 2010).

Der hier vorgestellte Kriterienkatalog REHBaR ist in der Grundlagenforschung zur Homöopathie das erste Instrument, das versucht, die Qualität von Publikationen zu verbessern. Mit Hilfe von 23 Kriterien (s. Tab 4-1) und einem ausführlichen Erläuterungskatalog mit Beispielen soll jeder Autor in die Lage versetzt werden, alle relevanten Informationen in einer Publikation wiederzugeben.

Eine Ursache für die oft lückenhafte Darstellung in vielen Publikationen liegt in den restriktiven Vorgaben vieler Zeitschriften an die Autoren, so dass sie detaillierte Beschreibungen und Datenübersichten aus Platzgründen nicht geben können. Die Verwendung einer Checkliste wie REBaHR soll helfen, den Fokus auf die erforderlichen und relevanten Informationen in einer Publikation zu lenken. Wenn z.B. die Methoden an einer anderen Stelle bereits publiziert wur-

den, reicht es aus auf diese Veröffentlichung hinzuweisen. Darüber hinaus sollten Autoren die Möglichkeit nutzen, als weitere Informationsquelle eine eigene Webseite zu benennen.

Im Bereich klinischer Studien wurde mit CONSORT bereits in den 1990er Jahren ein Richtmaß für eine standardisierte und qualitativ bessere Berichterstattung eingeführt. In allen großen medizinischen Zeitschriften wird auf die Einhaltung von CONSORT als Richtlinie für Autoren hingewiesen. Trotzdem hat diese konsequente Empfehlung nicht zu einer generellen Verbesserung der Berichterstattung geführt (Folkes et al. 2008; Brand 2009). Hier gibt es zum Teil erheblichen Nachholbedarf, sowohl innerhalb ganzer Fachrichtungen wie z.B. der Endokrinologie (Rios et al. 2008) als auch bezüglich der Qualität der Abstracts in medizinischen Zeitschriften (Berwanger et al. 2009). Eine Übersicht von Hopewell et al. aus dem Jahr 2010 untersucht die Qualität von mehr als 1000 Publikationen, die in Pubmed in den Jahren 2000 bis 2006 gelistet waren, also nachdem die CONSORT Richtlinie als Standard in den Journalen eingeführt wurde. Sie kommt zu dem Ergebnis, dass die Berichterstattung inzwischen zwar ausführlicher geworden ist, aber die Qualität weiter unter einem akzeptablen Zustand bleibt (Hopewell et al. 2010). Neben CONSORT hat auch STRICTA die Qualität der Publikationen (in diesem Falle von Akupunkturstudien) nicht in dem gewünschten und erwarteten Ausmaß verbessert (Prady et al. 2008). Ein Grund für diese bislang nicht adäquate Umsetzung der Empfehlungen liegt nach Vandenbroucke et al. an der Fülle der publizierten Richtlinien (Vandenbroucke 2009). Eine Fülle, die so unübersichtlich wurde, dass es bei EQUATOR neben der Verwaltung der Richtlinien inzwischen auch Schulungen, Workshops und Anleitungen zum Erstellen eigener Richtlinien gibt ([www.equator-network.org](http://www.equator-network.org)). Ein weiterer Grund für die ungenügende Umsetzung in der klinischen Forschung liegt in der nicht ausreichend konsequenten Begutachtung seitens der Journale, die zwar zum einen die Richtlinien vorgeben, zum anderen aber ihre Einhaltung nicht immer einfordern und kontrollieren (Hopewell et al. 2008).

Die vorliegende Arbeit schafft eine Grundlage für mehr Transparenz in Publikationen in der Grundlagenforschung zur Homöopathie. Eine Publikation ist die wichtigste und oft einzige Informationsquelle, um ein Experiment und dessen Ergebnisse klar und nachvollziehbar darzustellen. REHBaR ist der erste Katalog für den Bereich Grundlagenforschung zur Homöopathie und enthält Kriterien, die für alle Arten von Experimenten zu beschreiben sind. Damit soll das Instrument universell einsetzbar sein und nicht wie bei den klinischen Studien zu einer Vielzahl an spezialisierten Katalogen führen. Die Empfehlungen beziehen sich sowohl auf grundlegende wissenschaftliche Kriterien als auch auf spezielle Probleme in der Forschung zur Homöopathie.

Der Katalog ist in einem Team von Wissenschaftlern in einem Konsensverfahren erarbeitet worden. REHBaR und der Entwicklungsprozess sind bereits in zwei verschiedenen Journalen

publiziert worden: Homeopathy und ECAM (*Evidence based Complementary and Alternative Medicine*), letzteres ein frei zugängliches elektronisches Journal. Aus den Erfahrungen mit CONSORT wäre es im nächsten Schritt wichtig, dass die Journale, die Publikationen aus der Grundlagenforschung zur Homöopathie veröffentlichen, REHBaR als Richtlinie empfehlen und die Einhaltung der Kriterien einfordern. Auch das Expertenteam ist ein wichtiger Multiplikator, wenn alle Beteiligten in ihren Publikationen auf diese Richtlinie hinweisen.

## 5 Evaluation von REHBaR als Bewertungsinstrument für die Qualität von Publikationen

### 5.1 Hintergrund

Nachdem der Kriterienkatalog REHBaR fertig gestellt und publiziert war, sollte in einem zweiten Schritt geprüft werden, ob sich der Katalog als Bewertungsinstrument beim Lesen einer Publikation eignet. Die Experten des Delphi-Prozesses waren sich einig, dass für diese Form der Bewertung kein numerischer Score geeignet ist. Deswegen stellte sich die Frage, inwieweit REHBaR hilfreich sein kann, zu beurteilen, ob die publizierten Ergebnisse aus den verschiedenen Forschungsbereichen reliabel und valide sind und somit z.B. in eine qualitative Auswertung in Form eines systematischen Reviews eingeschlossen werden können.

Im ersten Schritt wurde der Katalog um drei Spalten erweitert, um ihn als REHBaR (Reviewing Experiments in Homeopathic Basic Research) nutzen zu können. Das Formular sieht vor, dass angekreuzt werden kann, ob die jeweilige Beschreibung gegeben ist („given“). Ist die Information zu dem Kriterium nicht gegeben oder nach Ansicht des Gutachters nicht ausreichend, bleibt die Spalte leer. Darüber hinaus soll angegeben werden, auf welcher Seite der Publikation die jeweilige Information zu finden ist („page“), und falls das Kriterium für die vorliegende Publikation nicht zutrifft, gibt es die Spalte „not applicable“. Dieser so abgeänderte Katalog wurde für die Pilotstudie mit den Publikationen gemeinsam an zwei Gutachter geschickt. Für die Hauptstudie wurden einige der Kriterien weiter unterteilt und erläutert sowie ein neues Kriterium (Nr. 24) mit aufgenommen. Der so modifizierte Katalog REHBaR enthält 39 Kriterien (s. Kap. 8.6)

### 5.2 Methodik

#### 5.2.1 Auswahl der Gutachter

Zwei Wissenschaftler konnten für diesen Beurteilungsprozess gewonnen werden, die nicht an der Entwicklung des Katalogs beteiligt waren, sich aber sowohl in der Grundlagenforschung gut auskennen als auch mit den Besonderheiten der Homöopathie vertraut sind. Dr. Otto Weingärtner ist in der Forschungsabteilung, Bereich medizinische Grundlagenforschung der pharmazeutischen Fabrik Dr. Reckeweg & Co. GmbH seit vielen Jahren tätig und publiziert auf dem Gebiet (Weingartner 2009). Der zweite Gutachter, Dr. Jürgen Clausen, ist Biologe und nach langjähriger Tätigkeit in der Grundlagenforschung in der Biologie/Botanik an der Universität Osnabrück (Clausen und Junge 2004) jetzt mit Fragestellungen zur Homöopathie befasst (van Wijk et al. 2009).

### 5.2.2 Auswahl der Publikationen

Es wurden nur Publikationen in die Beurteilung mit eingeschlossen, die nach der offiziellen Einführung der GLP-Richtlinien (good laboratory practice) 1981 publiziert worden sind. In den USA wurden diese Grundsätze im Jahr 1978 erstmalig veröffentlicht, in der OECD (Organisation for Economic Co-operation and Development) drei Jahre später als Standard in der Laborforschung fest gelegt. Im Jahr 1990 gründete sich das GLP-Panel, und gleichzeitig fanden die Laborrichtlinien Eingang ins Deutsche Recht, verankert im Chemikaliengesetz. Für die Einhaltung und Kontrolle ist die GLP-Bundesstelle im Bundesinstitut für Risikobewertung zuständig.

Als Grundlage für die Auswahl der Publikationen diente die Basic Research (BR) - Datenbank der Karl und Veronica Carstens-Stiftung (MySQL DB PHP Programm 5.10.2006). Die Datenbank verwaltet alle publizierten Experimente, die in der Grundlagenforschung zur Homöopathie durchgeführt wurden, mit Informationen zu verwendeten Versuchsmodellen, Arzneien und zugrunde liegenden Forschungsfragen (van Wijk und Albrecht 2007). Dabei wird die Datenbank in regelmäßigen Abständen aktualisiert (<http://www.carstens-stiftung.org>). Für den Zeitraum von 1/1981 bis 7/2009 waren in der Datenbank 920 Publikationen registriert. Diese lassen sich anhand des untersuchten Organismus („type of organism“) in folgende Kategorien einteilen:

**Animal (492), Fungus (26), Human (184), Microbial (40), Plant (101), None (74)**

Bei drei Publikationen ließ sich nicht feststellen, um welchen Organismus oder welche Art von Experiment es sich handelt. Diese wurden für das weitere Vorgehen ausgeschlossen. Unter der Kategorie „none“ sind die Experimente zusammen gefasst, die ohne Verwendung von biologischem Material durchgeführt wurden, z.B. physiko-chemische Versuche.

### 5.2.3 Pilotstudie

Die beiden Gutachter erhielten im August 2009 sieben Publikationen, die sie zur Prüfung des Instruments und Abklärung eventueller Verständnisfragen bewerten sollten. Dabei wurden per Zufallsziehung aus jeder Kategorie (s. oben) eine und aus dem „Animal“ Bereich zwei Publikationen ausgewählt. Die sieben Publikationen für die Pilotphase finden sich in Kap. 8.7.

Im September 2009 erfolgte ein Treffen mit den beiden Gutachtern, bei dem die Eignung von REHBaR diskutiert wurde. Der Katalog wurde daraufhin überarbeitet und weiter unterteilt. Dabei erschien es sinnvoll, die einzelnen Kriterien noch einmal zu gliedern, damit die jeweiligen Unterpunkte im Einzelnen angekreuzt werden konnten.

Außerdem wurde bei dem Treffen festgelegt, dass in der Hauptstudie erst einmal nur Publikationen bewertet werden sollten, die in wissenschaftlichen Zeitschriften veröffentlicht wurden. Für

Abstracts, Proceedings, Buchbeiträge und allgemeine Forschungsberichte sowie Dissertationen erschien das Instrument nicht geeignet.

### 5.2.4 Hauptstudie

Aus den 917 ausgewählten Publikationen der Basic Research Datenbank wurde per Zufallsziehung eine Auswahl an 50 Publikationen aus den verschiedenen Bereichen festgelegt. Dabei wurde dem Programm (MySQL) vorgegeben, wie viele Publikationen aus der Stichprobe zufällig ausgewählt werden sollten. Der Suchauftrag enthielt neben dem Bereich, in dem die Publikationen gesucht werden sollten, auch die Zeitangabe, aus der die Veröffentlichung stammen sollte. Von den insgesamt 917 Publikationen wurde somit eine Anzahl von Publikationen ermittelt, die im Verhältnis zur gesamten Anzahl der erfassten Experimente für den jeweiligen Bereich stand.

**Animal (27), Fungus (1), Human (10), Microbial (2), Plant (6), None (4)**

Von diesen 50 Publikationen wurden 33 in die Prüfung eingeschlossen, weil es sich hier um Publikationen in wissenschaftlichen Journalen handelte. Zu den 17 ausgeschlossenen Veröffentlichungen gehörten 5 Abstracts, 5 Buchbeiträge, 4 Dissertationen und 3 Mitteilungen/Kurzberichte.

Die Liste der von den beiden Gutachtern in der Hauptstudie zu bewertenden Publikationen (alphabetisch sortiert) findet sich in Kap. 8.7. Von den 33 zugeteilten Publikationen wurden 30 von beiden Gutachtern mithilfe des REHBaR – Katalogs in Kap. 8.6 bewertet.

### 5.2.5 Quantifizierung der Übereinstimmung – Kappa - Werte $\kappa$ und $\kappa_n$

In vielen Bereichen der Medizin und Psychologie ist es notwendig, ein Instrument zur Verfügung zu haben, anhand dessen die Einschätzung einer Situation/eines Zustands von verschiedenen Gutachtern sicher getroffen werden kann. In den Pflegewissenschaften ist die Zusammenarbeit verschiedener Professionen wie Pflegekräfte, Angehörige und Ärzte wichtig. Aus diesem Grund ist ein objektives Beurteilungsinstrument notwendig, um sich z.B. über den Zustand eines Patienten und damit auf die weitere Therapie zu einigen. Dieses Bewertungsinstrument sollte so zuverlässig sein, dass es erstens zu einer hohen Übereinstimmungsrate zweier Gutachter führt (Interrater Agreement), zweitens so sicher anwendbar sein, dass sich ein Gutachter in seinen Beurteilungen auch bei einer Neubewertung des gleichen Vorgangs nicht widerspricht (Intrarater Agreement) (Grouven et al. 2007).

Die einfachste Maßzahl für die Übereinstimmung zweier Gutachter ist die prozentuale Übereinstimmung. In der Tab. 5-1 ist sie als **relative Übereinstimmung** ( $p_0$ ) zu finden, die sich als

Quotient der gefundenen Übereinstimmung zur Gesamtzahl der bewerteten Publikationen berechnet.

Bei dieser einfachen Beschreibung wird allerdings außeracht gelassen, dass es eine zufällige Übereinstimmung geben kann. Um diese zufällige von der tatsächlichen Übereinstimmung zu differenzieren, hat Cohen einen Koeffizienten entwickelt: den so genannten **Kappa ( $\kappa$ )-Index** (Cohen 1960). Der  $\kappa$  – Wert ergibt sich für zwei Gutachter, in dem für vier Kombinationsmöglichkeiten die zu erwartenden Häufigkeiten berechnet werden. Dazu wird die jeweilige Zeilensumme mit der Spaltensumme multipliziert und dann durch die Gesamtzahl aller zu beurteilenden Kriterien dividiert (s. folgendes Beispiel).

<b>Abstimmung Gutachter 1 vs. 2</b>			
<b>Gutachter1</b>	<b>Gutachter 2</b>		
<b>Häufigkeit</b>	<b>0</b>	<b>X</b>	<b>Summe</b>
<b>0</b>	5	0	5
<b>x</b>	3	22	25
<b>Total</b>	8	22	30

Für das Beispiel aus der Tabelle errechnet sich demnach der  $\kappa$ -Wert wie folgt: Die beobachtete Übereinstimmung der beiden Gutachter ist die Summe aus „von beiden angekreuzt“ (22) plus „von beiden nicht angekreuzt“ (5), multipliziert mit der Gesamtzahl der Arbeiten, die mit dem gleichen Votum hätten belegt werden können (30) und bezogen auf alle Arbeiten (30):

$p_0 = 27/30 \times 30/30 = 0.9$  . Die zufällige Übereinstimmung errechnet sich aus dem Produkt der jeweiligen Spaltensumme und Zeilensumme, jeweils bezogen auf die Gesamtanzahl der bewerteten Arbeiten:

$$p_c = 5/30 \times 8/30 + 22/30 \times 25/30 = 0.65 .$$

Die tatsächliche Übereinstimmung (Cohens  $\kappa$ ) ergibt sich nun, in dem die zufällige von der beobachteten Übereinstimmung abgezogen wird, also:  $p_0 - p_c = 0.9 - 0.65 = 0.25$  und das Ergebnis dann zur Normierung durch  $(1 - \text{zufällige Übereinstimmung } p_c) = 1 - 0.65 = 0.35$  dividiert wird. Der für das Zahlenbeispiel errechnete Kappa ist:  **$\kappa = 0.71$** .

Die verwendete Formel lautet:

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

Mit Hilfe der SAS 9.2 Software (von SAS Inc. Cary North Carolina, USA) wurden die Kappa ( $\kappa$ ) -Werte mit den dazu gehörigen 95%-Konfidenzintervallen (95% CI) ermittelt und in Tab. 5-1 eingetragen.

Zusätzlich wurde der **korrigierte  $\kappa$  - Wert nach Brennan und Prediger** (Brennan und Prediger 1981) errechnet. Sie schlagen den Kappa-Wert  $\kappa_n$  vor, der  $p_c$  definiert als

$$p_c = \frac{1}{z},$$

wobei  $z$  die Anzahl der Kategorien (also in diesem Fall der Merkmale ja/nein) ist.

$$\kappa_n = \frac{p_0 - \frac{1}{z}}{1 - \frac{1}{z}}$$

Somit lautet  $\kappa_n$ :

Mit diesem korrigierten Wert gelingt es ein Problem zu lösen, das bei Cohens Kappa entsteht, wenn nur geringfügige Beurteilungsunterschiede auftreten, da der maximale Wert von Cohens Kappa von den Randverteilungen abhängt. Er wird umso geringer, je weiter sich die Randverteilungen von einer Gleichverteilung entfernen.  $\kappa_n$  führt in diesen Fällen dazu, dass sich hohe Übereinstimmungen auch sicher in hohen Kappa-Werten zeigen. Aus diesem Grund wird im Weiteren  $\kappa_n$  als Wert für die Übereinstimmung angenommen.

Die Klassifizierung der  $\kappa$ -Werte bezogen auf den Grad der Übereinstimmung erfolgt gemäß der am häufigsten zu findenden Einteilung nach Landis und Koch (Landis und Koch 1977):

$\kappa = 0,00$ bis $0,20$	sehr gering
$\kappa = 0,21$ bis $0,40$	gering
$\kappa = 0,41$ bis $0,60$	mittel
$\kappa = 0,61$ bis $0,80$	hoch
$\kappa = 0,81$ bis $1,00$	sehr hoch

Neben der Darstellung der  $\kappa$ - und  $\kappa_n$ - Werte empfehlen einige Autoren zur Verdeutlichung, die absolute Zahlen und das **95% Confidence Interval (CI)** anzugeben (Cicchetti und Feinstein 1990; Steinijans et al. 1997). Darüber hinaus dient die relative Häufigkeit der Übereinstimmung  $p_0$  (bei von Eye als „raw agreement“ bezeichnet) der zusätzlichen Veranschaulichung und wird in der Ergebnistabelle Tab. 5-1 mit aufgeführt. (von Eye 2006).



### **5.3 Ergebnisse der Evaluation – Messung der Reliabilität**

Die 23 Kriterien von REHBaR wurden in 39 Items (Unterpunkte) unterteilt. Für jeden dieser Items mussten die Gutachter ihr Votum abgeben, ob die Information in der Publikation ausreichend beschrieben wurde oder nicht. Bezogen auf den korrigierten Kappa-Wert  $\kappa_n$  kamen die beiden Gutachter bei 11 der 39 Punkte zu einer hohen bis sehr hohen Übereinstimmung ( $\kappa_n = 0.67 - 0.98$ ), bei 8 von 39 zu einer mittleren Übereinstimmung ( $\kappa_n = 0.47 - 0.60$ ). Bei 5 Items fand sich eine geringe ( $\kappa_n = 0.27$  und  $0.33$ ) und bei 15 Items eine sehr geringe Übereinstimmung ( $\kappa_n \leq 0.2$ ). Die Ergebnisse sind in Tab. 5-1 aufgeführt.

Tab. 5-1: Ergebnisse der Beurteilung von 30 Publikationen in der Grundlagenforschung zur Homöopathie mittels REHBaR. Die Interrater-Reliabilität von zwei Gutachtern wurde mittels Cohens  $\kappa$  und Brennan und Predigers  $\kappa_n$  ermittelt.  $\kappa$ -Werte werden gemeinsam mit dem Konfidenzintervall (CI) aufgelistet.  $\kappa_n$ -Werte  $\geq 0.60$  sind in fett gedruckt.

Nr.	Item	Descriptor	no	yes	Po	Kappa $\kappa_n$	Kappa $\kappa$ (95% CI)
1	Title	Title indicates the experimental model AND the intervention.	no	5	0	0.90	<b>0.80</b> 0.71 (0.41 to 1.00)
			yes	3	22		
2	Abstract	a) abstract provides an informative and balanced summary of what was done	no	2	0	0.93	<b>0.87</b> 0.63 (0.18 to 1.00)
			yes	2	26		
		b) abstract provides an informative and balanced summary of what was found	no	2	0	0.93	<b>0.87</b> 0.63 (0.18 to 1.00)
			yes	2	26		
3	Introduction	a) pertinent scientific literature and / or previous work on the subject is cited and gives a sufficient scientific background on the necessity of the presented research	no	4	2	0.80	<b>0.60</b> 0.44 (0.07 to 0.82)
			yes	4	20		
		b) experimental model is presented	no	2	2	0.53	0.07 0.02 (-0.24 to 0.27)
			yes	12	14		
		c) necessity of the investigation is explained	no	1	5	0.67	0.33 -0.04 (-0.38 to 0.30)
			yes	5	19		
		d) name of underlying homeopathic principle(s) (e.g. similia principle, potentisation, proving) and type of homeopathy (isopathy, classical vs complex homeopathy) is given	no	7	0	0.30	- 0.40 0.04 (-0.02 to 0.11)
			yes	21	2		
4	Objective/ Hypotheses	a) for confirmatory experiments: specific hypotheses and clearly defined primary outcome measure is given	no	14	0	0.53	0.07 0.12 (-0.04 to 0.28)
			yes	14	2		
		b) for exploratory experiments: hypotheses inducing the investigations are given.	no	5	18	0.37	- 0.27 0.04 (-0.13 to 0.21)
			yes	1	6		
5	Materials	Detailed description of all used materials a) devices (with ID),	no	11	0	0.40	- 0.20 0.04 (-0.04 to 0.12)
			yes	18	1		
		b) substances (non homeopathic) with charge-number (lot-number),	no	4	0	0.27	- 0.47 0.05 (-0.02 to 0.11)
			yes	22	4		

Nr.	Item	Descriptor	no		yes	P <sub>0</sub>	Kappa κ <sub>n</sub>	Kappa κ (95% CI)
5	Materials	c) instruments with sufficient information on origin (factory name, serial-number, town, country.)	no	13	0	0.60	0.20	0.27 (0.05 to 0.48)
			yes	12	5			
6	Materials (homeopathy specific)	a) manufacturer, Pharmacopoeia or production method for medications. If degree of potentization is given with D, C, CH or X, preparation method according to Hahnemann is defined.	no	4	3	0.87	<b>0.73</b>	0.59 (0.23 to 0.95)
			yes	1	22			
		b) potency or steps of dilution,	no	5	3	0.73	0.47	0.37 (0.01 to 0.72)
			yes	5	17			
		c) dilution method,	no	8	2	0.63	0.27	0.30 (-0.0 to 0.60)
			yes	9	11			
		d) starting point of dilution (e.g. mother tincture, D1, nosode).	no	15	6	0.80	<b>0.60</b>	0.60 (0.34 to 0.86)
			yes	0	9			
7	Homeopathic Controls	Description of how the control substance was prepared.	no	6	12	0.57	0.13	0.22 (-0.03 – 0.46)
			yes	1	11			
8	System Performance Controls	a) application of a test to demonstrate the stability of the test system (negative control),	no	10	15	0.47	- 0.07	0.09 (-0.11 to 0.3)
			yes	1	4			
		b) a test to demonstrate the reactivity of the system (positive control).	no	19	3	0.80	<b>0.60</b>	0.49 (0.14 to 0.84)
			yes	3	5			
9	Quality Control	Procedures and efforts used to enhance the quality and reliability of the experimental procedure	no	27	1	0.90	<b>0.80</b>	-0.05 (-0.11 to 0.02)
			yes	2	0			
10	Object of Investigation	Description for the particular system used in vivo, in vitro, biological, physical, biochemical, ....	no	2	11	0.60	0.20	0.10 (-0.14 to 0.35)
			yes	1	16			
11	Experimental Setup	a) detailed description of experimental conditions and procedures enabling the reader to replicate the experiment(s)	no	4	2	0.73	0.47	0.33 (-0.02 to 0.69)
			yes	6	18			
		b) number of treated units	no	7	3	0.40	- 0.20	-0.04 (-0.30 to 0.22)
			yes	15	5			
12	Replication	In case of internal replications: which materials have been reused; which have been changed.	no	25	2	0.87	<b>0.73</b>	0.26 (-0.26 to 0.78)
			yes	2	1			

Nr.	Item	Descriptor	no			yes	P <sub>0</sub>	Kappa κ <sub>n</sub>	Kappa κ (95% CI)
13	Parameters	Each parameter measured is named explicitly	no	1	8	0.67	0.33	0.02 <i>(-0.28 to 0.32)</i>	
			yes	2	19				
14	Intervention	Details on how and when substances or interventions for each group are actually administered.	no	5	14	0.43	- 0.13	-0.01 <i>(-0.28 to 0.26)</i>	
			yes	3	8				
15	Allocation	Method used to generate the group allocation including details (e.g. randomisation, blocking, stratification) has to be explained (term “randomised” is not sufficient).	no	27	0	0.93	<b>0.87</b>	0.47 <i>(-0.13 to 1.00)</i>	
			yes	2	1				
16	Blinding	Description if any procedure(s) or intervention(s) were concealed (term “blinded” is not sufficient).	no	23	5	0.80	<b>0.60</b>	0.38 <i>(-0.01 to 0.77)</i>	
			yes	2	1				
17	Statistical Methods	a) List of specific tests having been applied	no	9	4	0.73	0.47	0.46 <i>(0.14 to 0.78)</i>	
		yes	4	13					
		b) reason why the test(s) were applied	no	23	3	0.87	<b>0.73</b>	0.52 <i>(0.12 to 0.93)</i>	
		yes	1	3					
		c) information whether statistical tests were adjusted for multiplicity	no	27	0	0.97	<b>0.93</b>	0.78 <i>(0.37 to 1.00)</i>	
		yes	1	2					
18	Numbers analysed	a) Number of experiments with number of treated units (is this Number identical to Item 11b?)	no	11	2	0.47	- 0.07	0.02 <i>(-0.22 to 0.26)</i>	
		Yes	14	3					
		b) information on missing values	no	27	3	0.90	<b>0.80</b>	0	
		yes	0	0					
19	Data (descriptive)	Results are given in tables or figures (usually means and standard deviations, or medians and ranges are appropriate if parameters are scaled continuously. If scaled discretely, absolute frequencies and percentages are appropriate.)	no	1	3	0.80	<b>0.60</b>	0.13 <i>(-0.30 to 0.57)</i>	
			yes	3	23				
20	Data (inferential)	Appropriate measures of effect size (incl. uncertainty and probability) are given (mean differences between treatment and control (standard error or confidence intervals) and calculation of probability (p-values) are stated.))	no	6	18	0.33	- 0.33	-0.04 <i>(-0.25 to 0.17)</i>	
			yes	2	4				

Nr.	Item	Descriptor			no	yes	P <sub>0</sub>	Kappa $\kappa_n$	Kappa $\kappa$ (95% CI)
21	Interpretation	Interpretation of the results, taking into account study hypotheses or sources of potential bias or imprecision	no	1	10		0.60	0.20	-0.02 (-0.28 to 0.24)
			yes	2	17				
22	Evidence	General interpretation of results in the context of current evidence.	no	4	7		0.67	0.33	0.22 (-0.13 to 0.57)
			yes	3	16				
23	Experimental Model	Critical assessment of the chosen model or parameters.	no	11	4		0.67	0.33	0.33 (-0.00 to 0.67)
			yes	6	9				
24	Reference list	The list of cited literature is sufficient.	no	6	2		0.83	<b>0.67</b>	0.59 (0.27 to 0.91)
			yes	3	19				

In einem abschließenden Urteil sollten die beiden Gutachter entscheiden, welche der Publikationen sie aufgrund einer ausreichend guten Qualität in einen systematischen Review mit aufnehmen würden. Beide entschieden sich für die selben 26 der 30 Publikationen.

Eine **hohe bis sehr hohe Übereinstimmung** erzielten die beiden Gutachter bei den Items 1, 2a, 2b, 6a, 9, 12, 15 17b, 17c, 18b und 24. Das erste Item bezieht sich auf den Titel einer Publikation, ob z.B. das verwendete Untersuchungsmodell und die Behandlung detailliert beschrieben sind und der Titel aussagekräftig ist. Das war nach Meinung beider Gutachter bei 25 der 30 Publikationen gegeben ( $\kappa_n = 0.80$ ).

Beim Abstract musste beurteilt werden, ob beschrieben wird, was in dem Experiment genau durchgeführt wurde (Item 2a), und in einem zweiten Schritt, ob auch genaue Angaben gemacht wurden, zu welchem Ergebnis die Experimente kamen (Item 2b). Bei Item 2a und 2b gaben beide an, dass in 26 Publikationen die Informationen ausreichend waren, bei zwei dagegen nicht. Der Übereinstimmungskoeffizient für beide Items beträgt damit  $\kappa_n = 0.63$ .

Item 6a beinhaltet die Frage, ob die Informationen zur Herstellung der homöopathischen Testsubstanzen detailliert angegeben werden. Bei 22 Publikationen fanden beide Gutachter Angaben darüber, welche Ausgangssubstanz bzw. welche Ausgangspotenz für die homöopathischen Mittel verwendet wurde. Bei 4 Publikationen fehlten für beide Gutachter diese Angaben ( $\kappa_n = 0.73$ ).

Item 9 ist ein Kriterium, das abfragt, ob in der Publikation Angaben zur Qualitätssicherung beschrieben sind. Hier stimmten die Gutachter bei 27 der 30 Publikationen überein, dass die Information fehlte ( $\kappa_n = 0.80$ ).

Eine hohe Übereinstimmung ( $\kappa_n = 0.73$ ) erzielten die beiden Gutachter auch bei Item 12, der Frage, ob ausreichende Informationen zur Durchführung bei Replikationsversuchen gegeben wurden. Hier stimmten sie in 26 aller Fälle überein.

Ob die Versuche randomisiert durchgeführt wurden und dieser Prozess in der Publikation genauer erläutert wurde, fragt Item 15. Bei 27 der 30 Publikationen fanden beide Gutachter keine Angaben und kamen damit zu einer sehr hohen Übereinstimmung ( $\kappa_n = 0.87$ ).

Item 17 bezieht sich auf Angaben zur statistischen Auswertung der Experimente in der Publikation. Bei der Frage, ob Gründe für die Anwendung des jeweiligen Tests angegeben werden (Item 17b), kommen die Gutachter zu einer hohen Übereinstimmung ( $\kappa_n = 0.73$ ). Beim Unterpunkt 17c waren sich beide Gutachter in fast allen Publikationen (29 von 30) einig und erreichten eine sehr hohe Übereinstimmung ( $\kappa_n = 0.93$ ). Danach waren in 27 Publikationen keinerlei Angaben zu finden, ob eine Adjustierung bei multiplen Testen vorgenommen wurde. Bei zwei Publikationen fanden die Gutachter diese Angabe.

Bei Item 18b mussten sich die Gutachter entscheiden, ob in der Publikation Angaben gemacht wurden, wie mit fehlenden Werten bei der Auswertung rechnerisch verfahren wurde. Hier erzielten beide Gutachter eine hohe Übereinstimmung, da in 27 Fällen keine Information in der Publikation gegeben wurde ( $\kappa_n = 0.80$ ).

Das letzte Item, bei dem die beiden Gutachter eine hohe Übereinstimmung erreichten, fragt die Einschätzung ab, ob die Referenzliste an zitierter Literatur ausreichend ist (Item 24). Hier stimmten die beiden Gutachter bei 25 der 30 Publikationen in ihrer Einschätzung überein ( $\kappa_n = 0.67$ ).

**Mittlere Übereinstimmungen** erzielten die Gutachter bei den Items 3a, 6b, 6d, 8b, 11a, 16, 17a und 19. Item 3a fragt, ob in der Einleitung und Begründung für das Experiment hinreichend Literaturangaben gemacht werden und so anhand von Voruntersuchungen die Notwendigkeit des vorliegenden Experimentes begründet ist. Bei 24 Publikationen kamen beide Gutachter zur gleichen Bewertung ( $\kappa_n = 0.60$ ).

Bei Item 6b mussten die Gutachter entscheiden, ob die Informationen zur Herstellung der homöopathischen Mittel und hier insbesondere die einzelnen Potenzier- bzw. Verdünnungsschritte erläutert werden. Bei insgesamt 22 Publikationen kamen beide Gutachter zum gleichen Ergebnis ( $\kappa_n = 0.47$ ). In der Frage nach der Ausgangssubstanz der homöopathischen Mittelherstellung (Item 6d) kamen sie in 24 der 30 Publikationen zu einer mittleren Übereinstimmung ( $\kappa_n = 0.60$ ), dabei fanden sie bei der Hälfte der Publikationen (15) die Angaben, in 9 Publikationen nicht.

Bei Item 8b sollten die beiden Gutachter beurteilen, ob von einer Positivkontrolle berichtet wird oder nicht. Hier kreuzten sie bei nur fünf der 30 Publikationen ein „ja“ an, und fanden bei 19 Publikationen keinerlei Angaben zu dieser Kontrolle ( $\kappa_n = 0.60$ ).

Wie detailliert der Versuchsaufbau in der Publikation beschrieben ist, wird bei Item 11a abgefragt. Hier stimmten die beiden Gutachter bei 22 Einschätzungen überein und erzielten einen  $\kappa_n = 0.47$ .

Angaben in der Publikation zur Verblindung des Versuchsaufbaus bzw. des Versuchsleiters wird durch Item 16 erfasst. Hier fanden beide Gutachter bei 23 der Publikationen keine Angaben zur Verblindung ( $\kappa_n = 0.60$ ).

Beim ersten der drei Items (17a) erreichten die Gutachter bezüglich Angaben zur statistischen Auswertung eine mittlere Übereinstimmung ( $\kappa_n = 0.47$ ) und waren sich in der Einschätzung bei 22 der 30 Publikationen einig.

Item 19 bezieht sich auf die deskriptive Darstellung der Ergebnisse in einer Publikation in Form von Tabellen und Abbildungen. Hier gelangten die Gutachter zu einer mittleren Übereinstimmung ( $\kappa_n = 0.60$ ), d.h. sie beurteilten den Informationsgehalt in 24 Publikationen gleich.

Bei 20 Kriterien fiel die **Übereinstimmung** der beiden Gutachter **gering bis sehr gering** aus. Dazu gehören folgende Items: 3b, 3c, 3d, 4a, 4b, 5a, 5b, 5c, 6c, 7, 8, 10, 11b, 13, 14, 18a, 20, 21, 22 und 23.

Bei Item 3b wird die Frage gestellt, ob das Untersuchungsmodell genannt ist. Hier gingen die Einschätzungen in mehr als ein Drittel der Fälle (14 Publikationen) auseinander ( $\kappa_n = 0.07$ ).

Bei der Frage, ob das zugrunde liegende homöopathische Prinzip erwähnt wird (Item 3d), fielen die Beurteilungen sogar bei zwei Dritteln (21 Publikationen) gegensätzlich aus ( $\kappa_n = -0.40$ ).

Genauso uneins waren die Gutachter in der Frage nach der Beschreibung der Materialien (Item 5a bis c). Hier reichten Gutachter 1 die Daten bei allen drei Items in der Mehrzahl der Fälle aus (18, 22, 12), während Gutachter 2 diese Informationen als nicht gegeben oder unzureichend beurteilte.

Genau umgekehrt lag die Einschätzung bei der Frage nach der Negativkontrolle (Item 8a): hier kreuzte der zweite Gutachter in fast zwei Drittel der Fälle (19) ein „gegeben“ an, während Gutachter 1 dies nur in vier Publikationen für ausreichend erläutert befand.

Das gleiche Bild der auffallend unterschiedlichen Einschätzung der beiden Gutachter zeigt sich bei den Items 4 (Objectives /Hypotheses), 7 (Homeopathic Controls), 11b (Number of treated units), 14 (Intervention), 18 (Numbers analysed) und 20 (Data inferential).

## 5.4 Diskussion der Eignung des Kriterienkatalogs REHBaR

Zwei Gutachter sollten den Kriterienkatalog REHBaR evaluieren. Dazu wurden 30 Publikationen aus den verschiedenen Bereichen der Grundlagenforschung zur Homöopathie mit dem auf 39 Items (Unterpunkte) erweiterten Katalog bewertet. Ob die Gutachter zu der gleichen Einschätzung gelangten, wurde mit der relativen Übereinstimmung ( $p_0$ ), Cohens Kappa ( $\kappa$ ) und dem korrigierten Kappa nach Brennan und Prediger  $\kappa_n$  beschrieben. Bei Anwendung von  $\kappa_n$  kamen die Gutachter bei 19 Items zu einer mittleren bis hohen Übereinstimmung, bei 20 Items zu einer geringen bis sehr geringen Übereinstimmung. Im Folgenden wird diskutiert, warum mehr als die Hälfte der Items eine (sehr) geringe Übereinstimmung zeigen.

### 5.4.1 Eignung der $\kappa$ – Werte als Maß der Übereinstimmung

Cohens Kappa ist das am häufigsten eingesetzte Maß für die Übereinstimmung zweier Gutachter bei dichotomen Beurteilungsinstrumenten (zwei Antwortmöglichkeiten, z.B. „ja“ und „nein“) (Tooth und Ottenbacher 2004). Seit der Einführung von Cohens  $\kappa$  vor ca. 40 Jahren ist die Eignung von Cohens  $\kappa$  als statistische Methode immer wieder diskutiert worden (Kraemer et al. 2002). Darüber hinaus existieren eine Reihe von Ergänzungen und Korrekturformeln, um die Aussagekraft des  $\kappa$ -Wertes zu verbessern. Von Eye et al. schlussfolgern, dass der  $\kappa$ -Wert nicht als einzige Maßzahl für den Grad der Übereinstimmung angesehen werden darf (von Eye und von Eye 2008). Ein Hauptproblem bei der Berechnung von Cohens  $\kappa$  ist die Beeinflussung der Werte durch **Prävalenz** (Lantz und Nebenzahl 1996).

**Prävalenz** besagt in diesem Zusammenhang, mit welcher Häufigkeit ein bestimmtes Merkmal als positiv oder negativ eingeschätzt wird und welche Größe der entsprechende Faktor in der Berechnung z.B. des  $\kappa$ -Wertes annimmt. Rein rechnerisch ergeben sich die höchsten  $\kappa$ -Werte nur bei einer Gleichverteilung der Kreuze, d.h. wenn positive und negative Einschätzungen gleich häufig auftreten (Mayer et al. 2004; Wirtz und Kutschmann 2007).

Aus diesem Grund haben in den 1980er Jahren Brennan und Prediger den korrigierten Kappa  $\kappa_n$  vorgeschlagen, der in die Formel von  $\kappa$  die Zahl der Merkmalsträger ( $z$ ) – in diesem Fall „ja“ und „nein“ – mit einrechnet und so unregelmäßige Randverteilungen nivelliert. Als Beispiel zur Veranschaulichung sei die Berechnung von Item 9 gegeben: In 27 von 30 Publikationen waren sich die Gutachter einig, dass das Merkmal „nein“ gegeben war (Information nicht enthalten), ein „ja“ wurde in keinem Fall von beiden Gutachtern gleichzeitig angekreuzt. Diese Ungleichverteilung der Werte führt bei der Berechnung von  $\kappa$  trotz einer hohen relativen Häufigkeit ( $p_0$ ) zu einem negativen  $\kappa$ -Wert. Bei Brennan und Prediger hingegen spiegelt der  $\kappa_n$ -Wert mittels des Korrekturfaktor  $1/z$  (in diesem Falle  $= 1/2$ ) die hohe Übereinstimmung mit  $\kappa_n = 0.80$  wider.



#### 5.4.2 Eignung der Gutachter

Um eine valide Aussage über die Eignung eines Instrumentes machen zu können, sind nach Sadatsafavi et al. mehrere Gutachter empfehlenswert (Sadatsafavi et al. 2008). Eine höhere Power lässt sich erzielen, wenn z.B. drei oder vier Gutachter bewerten (Rigby 2000). Das war allerdings im Rahmen dieses Promotionsvorhabens nicht umzusetzen, da weitere Gutachter aufgrund des beträchtlichen Arbeitsaufwands nicht gewonnen werden konnten.

Bei nur zwei Gutachtern tritt ein Problem in den Vordergrund – der sogenannte **Response Bias** – welches vermutlich mit verantwortlich ist für die hohe Zahl an Items mit (sehr) geringer Übereinstimmung.

Mit Bias ist ein systematischer Fehler gemeint, der zu einer Verzerrung der Ergebnisse führt. Im vorliegenden Fall beschreibt er die Tendenz des einen Gutachters eine Kategorie eher mit „ja“ zu beurteilen, während der andere dazu neigt, diese eher mit „nein“ zu bewerten. Dieser so genannte „Response Bias“ kann z.B. auf einen unterschiedlichen Stand an Wissen oder Erfahrung der Gutachter zurückzuführen sein.

Die einzige Möglichkeit einen Response Bias als systematisch auftretenden Fehler zu minimieren, ist eine gute Schulung der beteiligten Gutachter (Tooth und Ottenbacher 2004). Ryan et al. schlussfolgern in ihrer Studie, dass nur durch eine gute Schulung und Einweisung der Gutachter eine akzeptable Interrater Reliabilität zu erhalten ist (Ryan et al. 1988); das legen auch die Ergebnisse von Myers et al. nahe (Myers et al. 2010). Neben der Schulung spielt die eigene Erfahrung des Gutachters eine große Rolle für die Reliabilität (Zhao et al. 2010) oder die Möglichkeit, sich neben dem Erhebungsinstrument andere Informationen für die Entscheidungsfindung anzueignen (Werner et al. 1983). In der Studie von Schuh et al. (Schuh et al. 2009) wird eine geringe Interrater Reliabilität auf unterschiedliche Voraussetzungen der Gutachter zurück geführt; hier wurden geschulte Experten mit Untersuchern verglichen, die erst vor Ort eine kurze Einweisung erhalten hatten.

Für die Evaluation von REHBaR wurden zwei ausgewiesene Experten als Gutachter ausgewählt, mit denen der Kriterienkatalog im Vorfeld intensiv besprochen und diskutiert wurde. Trotzdem scheinen unterschiedliche Vorkenntnisse und Erfahrungen zu der (sehr) geringen Übereinstimmung bei den Items 3b, 3d, 4a, 5a, 5b, 7, 8a, 11b, 14, 18 und 20 geführt zu haben.

#### 5.4.3 Eignung des Kriterienkatalogs

Ein anderer Grund für eine geringe Übereinstimmung unter Experten kann darin begründet liegen, dass einzelne Kriterien nicht eindeutig und präzise formuliert sind (Arimone et al. 2007). Eine fehlende Übereinstimmung kann somit an dem Fragestil liegen, z.B. wenn die vorgegebe-

nen Fragen eher allgemein gehalten sind oder Ansichten und eigene Erfahrungen die Beantwortung der Fragen mit beeinflussen (James et al. 1984).

Wenn man die 39 Items in zwei Kategorien einordnet, findet sich allerdings in der vorliegenden Arbeit das genaue Gegenteil dessen:

1. **unpräzise formulierte Kriterien**, die eine Beschreibung (description) oder eine Einschätzung (sufficient information) abfragen (Item 1, 2a, 2b, 3a, 3b, 3c, 7, 9, 10, 16, 17b, 21, 22, 23 und 24) und
2. **präzise formulierte Kriterien**, die eine genaue Zahl bzw. Information fordern (Item 3d, 4a, 4b, 5a, 5b, 5c, 6a, 6b, 6c, 6d, 8a, 8b, 11a, 11b, 12, 13, 14, 15, 17a, 17c, 18a, 18b, 19, 20).

Entgegen der ersten Erwartung kommen die beiden Gutachter bei den 15 unpräzisen Kriterien mit einem mittleren  $\kappa_n = 0.50 (\pm 0.29)$  zu einer deutlich besseren Übereinstimmung als bei den 24 präzise formulierten Kriterien mit  $\kappa_n = 0.25 (\pm 0.44)$  als Mittelwert. Das bedeutet, je ungenauer das Kriterium formuliert ist und je mehr die eigentliche Einschätzung bzw. Erfahrung des jeweiligen Gutachters gefragt ist, um so häufiger kommen beide zum gleichen Ergebnis. Dagegen fällt die Einschätzung um so uneinheitlicher aus, je präziser die Frage formuliert ist.

Dies lässt die Schlussfolgerung zu, dass sich das Raster von REHBaR bei seinen detailliert ausformulierten Kriterien nur schwer auf die oft ungenauen, nicht eindeutig zuzuordnenden Informationen in den evaluierten Publikationen anwenden ließ. Anders ausgedrückt: bei den präzise formulierten Kriterien fiel es den Gutachtern offensichtlich schwer, klare Antworten und Angaben in den Publikationen zu finden.

Ein Vergleich der 16 Items mit  $\kappa_n \geq 0.6$  (s. Tab. 5-1, S. 42 ff) bezüglich der positiven und negativen Übereinstimmungen stützt diese These. Bei den präzise formulierten Fragen stimmten die Gutachter nur bei 2 der 8 Items positiv überein und fanden 6 der Items nicht in den Publikationen. Es war also offenbar leichter, das Fehlen einer Information zu erkennen als sein Vorhandensein. Bei den unpräzisen Kriterien ist das Verhältnis umgekehrt: 5 Mal erachteten die Gutachter die Information als ausreichend (positive Übereinstimmung) und nur 3 Mal als unzureichend (negativ).

Zusammenfassend lässt sich aus der Evaluation von REHBaR weniger die Frage beantworten, ob bzw. inwieweit das Instrument für die Bewertung von Publikationen geeignet ist, als viel mehr, dass REHBaR nötig ist, damit Experimente aus der Grundlagenforschung zur Homöopathie zukünftig so publiziert werden, dass sie vollständig sowie eindeutig nachvollzogen und bewertet werden können.

Die Notwendigkeit, den Kriterienkatalog REHBaR als Standardinstrument für die Grundlagenforschung zur Homöopathie zu empfehlen, soll abschließend an folgendem Beispiel verdeutlicht werden: Bei der Befragung im ersten Teil der Arbeit (Kap.3) gaben 93 % (40 von 43) Teilnehmer an, dass Randomisierung und Verblindung in der Grundlagenforschung absolut notwendig bzw. sehr wichtig sei. In der Grundlagenforschung zur Homöopathie hingegen wird auch in einer ganz aktuellen Publikation von Chirumbolo et al. (2009) dezidiert nicht verblindet und randomisiert, ohne dafür eine Begründung zu geben. In den evaluierten Publikationen aus dem dritten Teil der Arbeit (Kap. 5) fanden die Gutachter nur bei einer von 30 Publikationen beschrieben, dass sie verblindet und randomisiert durchgeführt wurde. Mehr als zwei Drittel machten hierzu keinerlei Angaben (S. Tab. 5-1, S. 44).

Um die Qualität der Publikationen und damit vielleicht auch die Qualität in der Grundlagenforschung insgesamt zu verbessern, ist REHBaR entwickelt worden.

## 6 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde der Kriterienkatalog REHBaR (**R**eporting **E**xperiments in **H**omeopathic **B**asic **R**esearch) entwickelt. Dieser soll Wissenschaftlern, die eine experimentelle Arbeit aus dem Bereich der Grundlagenforschung zur Homöopathie publizieren wollen, als Anleitung dienen, damit die Experimente vollständig und transparent dargestellt werden. Diese Richtlinie soll dazu beitragen, dass die Berichterstattung standardisiert und transparenter wird und dient somit der Qualitätssicherung in der Grundlagenforschung zur Homöopathie.

Im ersten Teil der Arbeit wurden die 43 Antworten einer Befragung von 100 zufällig ausgewählten Forschungseinrichtungen außerhalb der Homöopathie vorgestellt. Die Befragung ergab, dass Randomisierung, Verblindung, interne Replikation sowie Positiv- und Negativkontrollen zu den Standards in der Grundlagenforschung gehören, wohingegen die externe Replikation von den meisten Forschern zwar als sehr wichtig erachtet, aber kaum durchgeführt wird.

Die Ergebnisse der Befragung dienten als Grundlage für die Erarbeitung des Kriterienkatalogs REHBaR. Gemeinsam mit einem Team von neun Wissenschaftlern aus dem Bereich der klinischen Forschung und der Grundlagenforschung zur Homöopathie wurde dieser in einem 5-stufigen Delphi-Prozess erarbeitet. REHBaR enthält insgesamt 23 Kriterien mit einem ausführlichen Erläuterungsteil inklusive Beispielen und soll als Anleitung für eine exakte und transparente Berichterstattung von Experimenten zu homöopathischen Fragestellungen dienen. Er ist gleichzeitig der erste Kriterienkatalog, der sämtliche Qualitätsstandards in der Grundlagenforschung zur Homöopathie enthält.

Inwieweit sich der Kriterienkatalog REHBaR als Evaluationsinstrument eignet, sollte im letzten Teil der Arbeit überprüft werden. Hierzu wurden die Kriterien des Katalogs von 23 auf 39 Unterpunkte weiter unterteilt. Zwei Gutachter bewerteten unabhängig voneinander 30 zufällig ausgewählte Publikationen aus verschiedenen Bereichen der Grundlagenforschung zur Homöopathie. Die Übereinstimmung wurde u.a. mit dem korrigierten Kappa-Wert ( $\kappa_n$ ) nach Brennan und Prediger ermittelt. Er zeigte bei 11 Kriterien eine hohe bis sehr hohe Übereinstimmung, bei 8 eine mittlere und bei 20 eine geringe bis sehr geringe Übereinstimmung. Ein Grund für die geringe Übereinstimmung lag an der oft ungenauen oder unvollständigen Berichterstattung in den Publikationen, die sich mit einem präzisen Instrument nicht sicher evaluieren lassen. So wurde bei unpräzise formulierten Kriterien durchschnittlich eine höhere Übereinstimmung zwischen den Gutachtern erzielt als bei präzise formulierte Kriterien. Damit zeigte die Evaluation von REHBaR weniger eine mangelnde Eignung des Katalogs als Bewertungsinstrument als vielmehr die Notwendigkeit, dass REHBaR als Richtlinie für eine vollständige und präzise Berichterstattung entwickelt wurde.

## 7 Literaturverzeichnis

1. Arimone, Y., Miremont-Salame, G., Haramburu, F., Molimard, M., Moore, N., Fourrier-Reglat, A., Begaud, B. (2007): Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions. *Br J Clin Pharmacol* 64 (4), 482-8.
2. Baumgartner, S. (2009): The state of basic research on homeopathy. In: C. Witt, Albrecht, H. (Eds.): *New directions in homeopathy research. Advice from an interdisciplinary conference*, S. 107-130. Essen: KVC-Verlag
3. Baumgartner, S., Heusser, P., Thurneysen, S. (1998): Methodological Standards and Problems in Preclinical Homoeopathic Potency Research. *Forsch Komplementarmed* 5 (1), 27-32.
4. Baumgartner, S., Shah, D., Schaller, J., Kampfer, U., Thurneysen, A., Heusser, P. (2008): Reproducibility of dwarf pea shoot growth stimulation by homeopathic potencies of gibberellic acid. *Complement Ther Med* 16 (4), 183-91.
5. Baumgartner, S., Thurneysen, A., Heusser, P. (2004): Growth stimulation of dwarf peas (*Pisum sativum* L.) through homeopathic potencies of plant growth substances. *Forsch Komplementarmed Klass Naturheilkd* 11 (5), 281-92.
6. Becker-Witt, C., Weissshuhn, T. E., Ludtke, R., Willich, S. N. (2003): Quality assessment of physical research in homeopathy. *J Altern Complement Med* 9 (1), 113-32.
7. Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D., Stroup, D. F. (1996): Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Jama* 276 (8), 637-9.
8. Berwanger, O., Ribeiro, R. A., Finkelsztejn, A., Watanabe, M., Suzumura, E. A., Duncan, B. B., Devereaux, P. J., Cook, D. (2009): The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol* 62 (4), 387-92.
9. Betti, L., Brizzi, M., Nani, D., Peruzzi, M. (1997): Effect of high dilutions of *Arsenicum album* on wheat seedlings from seed poisoned with the same substance. *Brit Hom J* 86(2), 86-89.
10. Betti, L., Trebbi, G., Majewsky, V., Scherr, C., Shah-Rossi, D., Jager, T., Baumgartner, S. (2009): Use of homeopathic preparations in phytopathological models and in field trials: a critical review. *Homeopathy* 98 (4), 244-66.
11. Bonamin, L. V., Endler, P. C. (2010): Animal models for studying homeopathy and high dilutions: conceptual critical review. *Homeopathy* 99 (1), 37-50.
12. Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. (2003): Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Acad Radiol* 10 (6), 664-9.
13. Boutron, I., Tubach, F., Giraudeau, B., Ravaud, P. (2003): Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *Jama* 290 (8), 1062-70.
14. Brack, A., Strube, J., Stolz, P., Decker, H. (2003): Effects of ultrahigh dilutions of 3,5-dichlorophenol on the luminescence of the bacterium *Vibrio fischeri*. *Biochim Biophys Acta* 1621 (3), 253-60.

15. Brand, R. A. (2009): Standards of reporting: the CONSORT, QUORUM, and STROBE guidelines. *Clin Orthop Relat Res* 467 (6), 1393-4.
16. Brennan, R. L., Prediger, D. J. (1981): Coefficient  $\kappa$ : Some uses, misuses, and alternatives. *Educational and Psychological Measurement*. 41, 687-699.
17. Brizzi, M., Lazzarato, L., Nani, D., Borghini, F., Peruzzi, M., Betti, L. (2005): A biostatistical insight into the As(2)O(3) high dilution effects on the rate and variability of wheat seedling growth. *Forsch Komplementarmed Klass Naturheilkd* 12 (5), 277-83.
18. Cazin, J. C., Cazin, M., Gaborit, J. L., Chaoui, A., Boiron, J., Belon, P., Cherruault, Y., Papapanayotou, C. (1987): A Study of the Effect of Decimal and Centesimal Dilutions of Arsenic on the Retention and Mobilization of Arsenic in the Rat. *Human Toxicol* 6, 315-320.
19. Cesar, B., Abud, A. P., de Oliveira, C. C., Cardoso, F., Gremski, W., Gabardo, J., Buchi Dde, F. (2008): Activation of mononuclear bone marrow cells treated in vitro with a complex homeopathic medication. *Micron* 39 (4), 461-70.
20. Chirumbolo, S., Brizzi, M., Ortolani, R., Vella, A., Bellavite, P. (2009): Inhibition of CD203c membrane up-regulation in human basophils by high dilutions of histamine: a controlled replication study. *Inflamm Res* 58 (11), 755-64.
21. Cicchetti, D. V., Feinstein, A. R. (1990): High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 43 (6), 551-8.
22. Clausen, J., Junge, W. (2004): Detection of an intermediate of photosynthetic water oxidation. *Nature* 430 (6998), 480-3.
23. Cohen, J. A. (1960): A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46.
24. Davidoff, F., Batalden, P., Stevens, D., Ogrinc, G., Mooney, S. E. (2009): Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *Bmj* 338, a3152.
25. Dean, M. E., Coulter, M. K., Fisher, P., Jobst, K. A., Walach, H. (2007): Reporting data on homeopathic treatments (RedHot): a supplement to CONSORT. *J Altern Complement Med* 13 (1), 19-23.
26. Elia, V. E., L. Montanino, M., Napoli, E., Niccoli, M., Nonatelli, L. (2007): Conductometric studies of the serially diluted and agitated solutions on an anomalous effect that depends on the dilution process. *J. Molecular Liquids* 135, 158-165.
27. Endler, P., Thieves, K., Reich, C., Matthiessen, P., Bonamin, L., Scherr, C., Baumgartner, S. (2009): Repetitions of fundamental research models for homeopathically prepared dilutions beyond 10(-23): a bibliometric study. *Homeopathy* 99 (1), 25-36.
28. Flower, A., Lewith, G. T., Little, P. (2007): Seeking an oracle: using the Delphi process to develop practice guidelines for the treatment of endometriosis with Chinese herbal medicine. *J Altern Complement Med* 13 (9), 969-76.
29. Folkes, A., Urquhart, R., Grunfeld, E. (2008): Are leading medical journals following their own policies on CONSORT reporting? *Contemp Clin Trials* 29 (6), 843-6.
30. Gagnier, J. J., Boon, H., Rochon, P., Moher, D., Barnes, J., Bombardier, C. (2006): Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 144 (5), 364-7.

31. Grouven, U., Bender, R., Ziegler, A., Lange, S. (2007): [The kappa coefficient]. *Dtsch Med Wochenschr* 132 Suppl 1, e65-8.
32. Hopewell, S., Altman, D. G., Moher, D., Schulz, K. F. (2008): Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials* 9, 20.
33. Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D. G., Schulz, K. F. (2008): CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5 (1), e20.
34. Hopewell, S., Dutton, S., Yu, L. M., Chan, A. W., Altman, D. G. (2010): The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *Bmj* 340, c723.
35. Hornung, J., Linde, K. (1991): Guidelines for the Exact Description of the Preparation and Mode of Application of Serial Dilutions and Potencies in Publications on Ultra Low Dose Effects and Homeopathic Research - a proposal-. *The Berlin Journal on Research in Homeopathy* 1 (2), 121-123.
36. Ives, J. A., Moffett, J. R., Arun, P., Lam, D., Todorov, T. I., Brothers, A. B., Anick, D. J., Centeno, J., Namboodiri, M. A., Jonas, W. B. (2010): Enzyme stabilization by glass-derived silicates in glass-exposed aqueous solutions. *Homeopathy* 99 (1), 15-24.
37. James, L. R., Demaree, R. G., Wolf, G. (1984): Estimating Within-Group Interrater Reliability with and without Response Bias. *Journal of Applied Psychology* 69 (1), 85-98.
38. Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hrobjartson, A., Roberts, C., Shoukri, M., Streiner, D. L. (2010): Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* in press.
39. Kraemer, H. C., Periyakoil, V. S., Noda, A. (2002): Tutorial in Biostatistics - Kappa coefficients in medical research. *Statist. Med.* 21, 2109-2129.
40. Lahnstein, L., Binder, M., Thurneysen, A., Frei-Erb, M., Betti, L., Peruzzi, M., Heusser, P., Baumgartner, S. (2009): Isopathic treatment effects of Arsenicum album 45x on wheat seedling growth--further reproduction trials. *Homeopathy* 98 (4), 198-207.
41. Landis, J. R., Koch, G. G. (1977): The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159-74.
42. Lantz, C. A., Nebenzahl, E. (1996): Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J Clin Epidemiol* 49 (4), 431-4.
43. Linde, K., Clausius, N., Ramirez, G., Melchart, D., Eitel, F., Hedges, L. V., Jonas, W. B. (1997): Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 350 (9081), 834-43.
44. Linde, K., Jonas, W. B., Melchart, D., Worku, F., Wagner, H., Eitel, F. (1994): Critical review and meta-analysis of serial agitated dilutions in experimental toxicology. *Hum Exp Toxicol* 13 (7), 481-92.
45. Linde, K. M., D; Jonas, W.B.; Hornung, J. (1994): Ways to enhance the quality and acceptance of clinical and laboratory studies in homeopathy. *Brit Hom J* 83 (1), 3-7.
46. Linstone, H., Turoff, M. (2002). *The Delphi Method: Techniques and Applications*. Online-Publikation, <http://www.is.njit.edu/pubs/delphibook/>.

47. Little, J., Higgins, J. P., Ioannidis, J. P., Moher, D., Gagnon, F., von Elm, E., Khoury, M. J., Cohen, B., Davey-Smith, G., Grimshaw, J., Scheet, P., Gwinn, M., Williamson, R. E., Zou, G. Y., Hutchings, K., Johnson, C. Y., Tait, V., Wiens, M., Golding, J., van Duijn, C., McLaughlin, J., Paterson, A., Wells, G., Fortier, I., Freedman, M., Zecevic, M., King, R., Infante-Rivard, C., Stewart, A., Birkett, N. (2009): STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med* 6 (2), e22.
48. Ludtke, R., Rutten, A. L. (2008): The conclusions on the effectiveness of homeopathy highly depend on the set of analyzed trials. *J Clin Epidemiol* 61 (12), 1197-204.
49. MacPherson, H., White, A., Cummings, M., Jobst, K., Rose, K., Niemtzow, R. (2001): Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. *Complement Ther Med* 9 (4), 246-9.
50. Majewsky, V., Arlt, S., Shah, D., Scherr, C., Jager, T., Betti, L., Trebbi, G., Bonamin, L., Klocke, P., Baumgartner, S. (2009): Use of homeopathic preparations in experimental studies with healthy plants. *Homeopathy* 98 (4), 228-43.
51. Mayer, H., Nonn, C., Osterbrink, J., Evers, G. C. (2004): [Quality criteria of assessment scales--Cohen's kappa as measure of interrater reliability (1)]. *Pflege* 17 (1), 36-46.
52. Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F. (1999): Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses. Lancet* 354 (9193), 1896-900.
53. Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., Walsh, S. (1995): Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 16 (1), 62-73.
54. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009): Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J Clin Epidemiol*.
55. Moher, D., Schulz, K. F., Altman, D. (2005): The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. *Explore (NY)* 1 (1), 40-5.
56. Moher, D., Schulz, K. F., Altman, D. G. (2001): The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *J Am Podiatr Med Assoc* 91 (8), 437-42.
57. Myers, J., Gardiner, K., Harris, K., Lilien, T., Bennett, M., Chow, E., Selby, D., Zhang, L. (2010): Evaluating correlation and interrater reliability for four performance scales in the palliative care setting. *J Pain Symptom Manage* 39 (2), 250-8.
58. Nani, D., Brizzi, M., Lazzarato, L., Betti, L. (2007): The role of variability in evaluating ultra high dilution effects: considerations based on plant model experiments. *Forsch Komplement Med* (2006) 14 (5), 301-5.
59. Prady, S. L., Macpherson, H. (2007): Assessing the utility of the standards for reporting trials of acupuncture (STRICTA): a survey of authors. *J Altern Complement Med* 13 (9), 939-43.
60. Prady, S. L., Richmond, S. J., Morton, V. M., Macpherson, H. (2008): A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE* 3 (2), e1577.



61. Rigby, A. S. (2000). "Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient." *Disabil Rehabil* **22**(8): 339-44.
62. Rios, L. P., Odueyungbo, A., Moitri, M. O., Rahman, M. O., Thabane, L. (2008): Quality of reporting of randomized controlled trials in general endocrinology literature. *J Clin Endocrinol Metab* **93** (10), 3810-6.
63. Ryan, J. W., Phillips, C. Y., Prescott, P. A. (1988): Interrater reliability: the underdeveloped role of rater training. *Appl Nurs Res* **1** (3), 148-50.
64. Sadatsafavi, M., Najafzadeh, M., Lynd, L., Marra, C. (2008): Reliability studies of diagnostic tests are not using enough observers for robust estimation of interobserver agreement: a simulation study. *J Clin Epidemiol* **61** (7), 722-7.
65. Scherr, C., Simon, M., Spranger, J., Baumgartner, S. (2009): Effects of potentised substances on growth rate of the water plant *Lemna gibba* L. *Complement Ther Med* **17** (2), 63-70.
66. Schuh, L. A., London, Z., Neel, R., Brock, C., Kissela, B. M., Schultz, L., Gelb, D. J. (2009): Education research: Bias and poor interrater reliability in evaluating the neurology clinical skills examination. *Neurology* **73** (11), 904-8.
67. Schulte, J. (1994): About Quality and standards in Ultra High Dilution Research and Research on Homeopathy. In: P. C. S. Endler, J. (Eds.): *Ultra High Dilution - Physiology and Physics*, S. 171-176. Dordrecht: Kluwer Academic Publishers
68. Shang, A., Huwiler-Muntener, K., Nartey, L., Juni, P., Dorig, S., Sterne, J. A., Pewsner, D., Egger, M. (2005): Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *Lancet* **366** (9487), 726-32.
69. Simera, I., Moher, D., Hoey, J., Schulz, K. F., Altman, D. G. (2009): The EQUATOR Network and reporting guidelines: Helping to achieve high standards in reporting health research studies. *Maturitas* **63** (1), 4-6.
70. Steinijans, V. W., Diletti, E., Bomches, B., Greis, C., Solleder, P. (1997): Interobserver agreement: Cohen's kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J Clin Pharmacol Ther* **35** (3), 93-5.
71. Stock-Schroer, B., Albrecht, H., Betti, L., Dobos, G., Endler, C., Linde, K., Ludtke, R., Musial, F., van Wijk, R., Witt, C., Baumgartner, S. (2009): Reporting Experiments in Homeopathic Basic Research--Description of the Checklist Development. *Evid Based Complement Alternat Med*.
72. Stock-Schroer, B., Albrecht, H., Betti, L., Endler, P. C., Linde, K., Ludtke, R., Musial, F., van Wijk, R., Witt, C., Baumgartner, S. (2009): Reporting experiments in homeopathic basic research (REHBaR)--a detailed guideline for authors. *Homeopathy* **98** (4), 287-98.
73. Taylor, C. F., Field, D., Sansone, S. A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P. A., Bogue, M., Booth, T., Brazma, A., Brinkman, R. R., Michael Clark, A., Deutsch, E. W., Fiehn, O., Fostel, J., Ghazal, P., Gibson, F., Gray, T., Grimes, G., Hancock, J. M., Hardy, N. W., Hermjakob, H., Julian, R. K., Jr., Kane, M., Kettner, C., Kinsinger, C., Kolker, E., Kuiper, M., Le Novere, N., Leebens-Mack, J., Lewis, S. E., Lord, P., Mallon, A. M., Marthandan, N., Masuya, H., McNally, R., Mehrle, A., Morrison, N., Orchard, S., Quackenbush, J., Reecy, J. M., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Rosenfelder, H., Santoyo-Lopez, J., Scheuermann, R. H., Schober, D., Smith, B., Snape, J., Stoeckert, C. J., Jr., Tipton, K., Sterk, P., Untergasser, A., Vandesompele, J.,

- Wiemann, S. (2008): Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol* 26 (8), 889-96.
74. Tooth, L. R., Ottenbacher, K. J. (2004): The kappa statistic in rehabilitation research: an examination. *Arch Phys Med Rehabil* 85 (8), 1371-6.
75. Turpin, D. L. (2005): CONSORT and QUOROM guidelines for reporting randomized clinical trials and systematic reviews. *Am J Orthod Dentofacial Orthop* 128 (6), 681-5; discussion 686.
76. van Wijk, R., Albrecht, H. (2007): Classification of systems and methods used in biological basic research on homeopathy. *Homeopathy* 96 (4), 247-51.
77. van Wijk, R., Albrecht, H. (2007): Proving and therapeutic experiments in the HomBRex basic homeopathy research database. *Homeopathy* 96 (4), 252-7.
78. van Wijk, R., Clausen, J., Albrecht, H. (2009): The rat in basic therapeutic research in homeopathy. *Homeopathy* 98 (4), 280-6.
79. Vandembroucke, J. P. (2009): STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol* 62 (6), 594-6.
80. Vandembroucke, J. P., von Elm, E., Altman, D. G., Gotzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., Egger, M. (2007): Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 147 (8), W163-94.
81. Velkers, F. C., Loo, A. J. H. t., Madin, F., Eck, J. H. H. v. (2005): Isopathic and pluralist homeopathic treatment of commercial broilers with experimentally induced colibacillosis. *Res Vet Science* 78(1), 77-83.
82. Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., Knipschild, P. G. (1998): The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 51 (12), 1235-41.
83. Vickers, A. J. (1999): Independent replication of pre-clinical research in homeopathy: a systematic review. *Forsch Komplementarmed* 6 (6), 311-20.
84. von Eye, A. (2006): An Alternative to Cohen's  $\kappa$ . *European Psychologist* 11 (1), 12-24.
85. von Eye, A., von Eye, M. (2008): On the Marginal Dependency of Cohen's  $\kappa$ . *European Psychologist* 13 (4), 305-315.
86. Walchli, C., Baumgartner, S., Bastide, M. (2006): Effect of low doses and high homeopathic potencies in normal and cancerous human lymphocytes: an in vitro isopathic study. *J Altern Complement Med* 12 (5), 421-7.
87. Weber, S., Endler, P. C., Welles, S. U., Suanjak-Traidl, E., Scherer-Pongratz, W., Frass, M., Spranger, H., Peithner, G., Lothaller, H. (2008): The effect of homeopathically prepared thyroxine on highland frogs: influence of electromagnetic fields. *Homeopathy* 97 (1), 3-9.
88. Weingartner, O. (2009): A formal approach to the problem of reproducing experimental effects with homeopathic potencies. *J Altern Complement Med* 15 (3), 287-91.
89. Werner, P. D., Rose, T. L., Yesavage, J. A. (1983): Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness. *J Consult Clin Psychol* 51 (6), 815-25.

90. Wiegant, F. A., Souren, J. E., van Wijk, R. (1999): Stimulation of survival capacity in heat shocked cells by subsequent exposure to minute amounts of chemical stressors; role of similarity in hsp-inducing effects. *Hum Exp Toxicol* 18 (7), 460-70.
91. Wiegant, F. A. C., Souren, J. E. M., Rijn, J. v., Wijk, R. v. (1994): Stressor-specific Induction of Heat Shock Proteins in Rat Hepatoma Cells. *Toxicology* 94, 143-159.
92. Wijk, R. v., Bosman, S., Wijk, E. P. v. (2006): Thermoluminescence in Ultra-High Dilution Research. *J Altern Complement Med* 12 (5), 437-443.
93. Wirtz, M., Kutschmann, M. (2007): Analyzing interrater agreement for categorical data using Cohen's kappa and alternative coefficients. *Rehabilitation (Stuttg)* 46 (6), 370-7.
94. Witt, C. (2000): *Physikalische Untersuchung homöopathischer Hochpotenzen*. Essen: KVC-Verlag; 1-82
95. Witt, C., Ludtke, R., Weissshuhn, T. E., Willich, S. N. (2005): High homeopathic potencies are different from potentized solvent when investigated with the RE-DEM technology. *Forsch Komplementarmed Klass Naturheilkd* 12 (1), 6-13.
96. Witt, C. M., Bluth, M., Albrecht, H., Weissshuhn, T. E., Baumgartner, S., Willich, S. N. (2007): The in vitro evidence for an effect of high homeopathic potencies--a systematic review of the literature. *Complement Ther Med* 15 (2), 128-38.
97. Witt, C. M., Bluth, M., Hinderlich, S., Albrecht, H., Ludtke, R., Weissshuhn, T. E., Willich, S. N. (2006): Does potentized HgCl<sub>2</sub> (Mercurius corrosivus) affect the activity of diastase and alpha-amylase? *J Altern Complement Med* 12 (4), 359-65.
98. Wolf, U., Wolf, M., Heusser, P., Thurneysen, A., Baumgartner, S. (2009): Homeopathic Preparations of Quartz, Sulfur and Copper Sulfate Assessed by UV-Spectroscopy. *Evid Based Complement Alternat Med* in press.
99. Zhao, H., Collier, J. M., Quah, D. M., Purvis, T., Bernhardt, J. (2010): The modified Rankin Scale in acute stroke has good inter-rater-reliability but questionable validity. *Cerebrovasc Dis* 29 (2), 188-93.
100. Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., Moher, D. (2008): Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *Bmj* 337, a2390.

## 8 Anhang

### 8.1 Abkürzungsverzeichnis

Abb.	Abbildung
BEPEV	Beurteilungs-Score für Physikalisch –Experimentelle Versuchsreihen
Kap.	Kapitel
RCT	Randomised Clinical Trials. Randomisierte Kontrollierte klinische Studie
REHBaR	Reporting Experiments in Homeopathic Basic Research
SAPEH	Score for Assessment of Physical Experiments on Homeopathy
Tab.	Tabelle
UHD	Ultra High Dilution: Verdünnungen von Stoffen oberhalb der Avogadro Konstante (sie hat den Wert: $N_A = 6,022 \cdot 10^{23} \text{ mol}^{-1}$ ). Demnach kann kein Molekül der Ausgangssubstanz mehr enthalten sein.

## 8.2 Fragebogen zu ausgewählten Standards in der Laborpraxis

(Bitte kreuzen Sie immer nur eine Antwortmöglichkeit an)

### **1. Randomisierung** der Proben (zufällige Anordnung der Versuchseinheiten)

☐ halte ich für absolut notwendig / mache ich immer

☐ halte ich für sehr wichtig / mache ich aber nicht immer, weil

---

---

☐ halte ich für verzichtbar, weil

---

---

### **2. Verblindung** (Kontrolle und Verum werden kodiert, bei der Versuchsdurchführung nicht erkenntlich)

☐ halte ich für absolut notwendig / mache ich immer

☐ halte ich für sehr wichtig / mache ich aber nicht immer, weil

---

---

☐ halte ich für verzichtbar, weil

---

---

### **3. Positiv und Negativkontrollen**

☐ halte ich für absolut notwendig / mache ich immer

☐ halte ich für sehr wichtig / mache ich aber nur bei

a) ☐ Positivkontrolle (zur Kontrolle der Reagibilität des Systems)

b) ☐ Negativkontrolle (zur Kontrolle der Stabilität des Versuchssystems)

☐ halte ich für verzichtbar, weil

---

---

### **4. interne Replikation** der Versuche zu einem späteren Zeitpunkt mit neuen Versuchsmaterialien:

☐ halte ich für absolut notwendig / mache ich immer

☐ halte ich für sehr wichtig / mache ich aber nicht immer, weil

---

---

☐ halte ich für verzichtbar, weil

---

---

**5. externe Replikation** der Versuche (in einem anderen Labor):

☐ halte ich für absolut notwendig / mache ich immer

☐ halte ich für sehr wichtig / mache ich aber nicht immer, weil

---

---

☐ halte ich für verzichtbar, weil

---

---

**6. für die statistische Planung und Auswertung**

☐ ziehe ich immer einen Statistiker zu Rate

☐ kontaktiere ich den Statistiker nur in Ausnahmefällen

☐ brauch ich keinen Statistiker, weil ich die Auswertung selber machen kann

**7. Anmerkungen:**

### 8.3 Übersicht der angeschriebenen Institute

<i>Pflanzenphysiologie</i>	
Institut für Biologie II RWTH Aachen	Aachen
Lehrstuhl Pflanzenphysiologie, Universität Bayreuth	Bayreuth
Arbeitsgruppe Pflanzenphysiologie, Institut für Biologie, FU Berlin	Berlin
Lehrstuhl für Pflanzenphysiologie, Ruhr-Universität Bochum	Bochum
Nees-Institut für Biodiversität der Pflanzen	Bonn
FB 2 Biologie - Botanik Molekulare Pflanzenphysiologie, Universität Bremen	Bremen
Institut für Ökologische Pflanzenphysiologie, Universität Düsseldorf	Düsseldorf
Lehrstuhl für Molekulare Pflanzenphysiologie, Universität Erlangen-Nürnberg	Erlangen
FB BioGeo - Botanik, Universität Duisburg-Essen	Essen
Institut für Biologie 2 und Pflanzenbiotechnologie, Universität Freiburg	Freiburg
Lehrstuhl für Entwicklungsbiologie der Pflanzen, Universität Freiburg	Freiburg
Institut für Gartenbau	Freiburg
FB Biologie, Institut für Pflanzenphysiologie, Universität Gießen	Gießen
Institut für Biologie, Universität Potsdam	Golm
Max-Planck-Institut für Molekulare Pflanzenphysiologie	Golm
FB Biologie, Institut für Pflanzenphysiologie, Universität Halle-Wittenberg	Halle
Biozentrum Klein Flottbek Allgemeine und Angewandte Pflanzenphysiologie	Hamburg
Institut für Allgemeine Botanik und Pflanzenphysiologie, Universität Jena	Jena
Lehrstuhl für Pflanzenphysiologie und Evolutionsbiologie, Universität Kassel	Kassel
Botanisches Institut Abt. Ökophysiologie der Pflanzen, Universität Kiel	Kiel
Botanisches Institut, Universität Köln	Köln
Institut für allgemeine Botanik, Universität Mainz	Mainz
Fb. 17 - Biologie, Universität Marburg	Marburg
FB Biologie, Institut für Pflanzenphysiologie, Universität Osnabrück	Osnabrück

Fakultät Agrarwissenschaften und Landschaftsarchitektur- Pflanzenbau, Fachhochschule Osnabrück	Osnabrück
Institut für Biowissenschaften Abteilung Pflanzenphysiologie, Universität Rostock	Rostock
Institut für Physiologie und Biotechnologie der Pflanzen, Universität Hohenheim	Stuttgart
Lehrstuhl für Pflanzenphysiologie, Universität Tübingen	Tübingen
Zentrum für Molekularbiologie der Pflanzen, Universität Tübingen	Tübingen
Lehrstuhl für Botanik I - Pflanzenphysiologie und Biophysik, Universität Würzburg	Würzburg
<i>Tierphysiologie</i>	
Institut für Veterinär-Physiologie, FU-Berlin	Berlin
Lehrstuhl für Tierphysiologie, HU Berlin	Berlin
Lehrstuhl für Tierphysiologie, Universität Bochum	Bochum
Zoologisches Institut - Abt. Tierphysiologie, Universität Braunschweig	Braunschweig
Johannes von Lintig Institut für Biologie/Neurobiologie, Tierphysiologie, Universität Freiburg	Freiburg i.Br
Institut für Tierphysiologie, Universität Gießen	Gießen
Institut für Tierphysiologie, Universität Gießen	Gießen
Institut für Tierphysiologie und Tierernährung, Universität Göttingen	Göttingen
Institut für Allgemeine Zoologie und Tierphysiologie, Universität Jena	Jena
Zoologisches Institut, Universität Köln	Köln
Veterinärmedizinische Fakultät, Universität Leipzig	Leipzig
Institut für Zoologie, Universität Mainz	Mainz
Institut für Tierphysiologie, Universität Marburg	Marburg
Institut für Physiologie, Physiologische Chemie und Tierernährung, LMU München	München
Institut für Tierphysiologie, Universität Münster	Münster
Institut für Tierphysiologie, Universität Osnabrück	Osnabrück
Institut für Zoologie, Universität Regensburg	Regensburg
Institut für Zellbiologie und Biosystemtechnik Abt. Tierphysiologie, Universität Rostock	Rostock



FR 8.3, Zoologie / Physiologie, Universität des Saarlandes	Saarbrücken
Biologisches Institut, Abt. Tierphysiologie, Universität Stuttgart	Stuttgart
Lehrstuhl für Tierphysiologie, Universität Tübingen	Tübingen
<i>Pharmakologie und Toxikologie</i>	
Leibniz-Institut für Molekulare Pharmakologie, Campus Berlin-Buch	Berlin
FB Veterinärmedizin Institut für Pharmakologie und Toxikologie, FU Berlin	Berlin
Institut für Pharmakologie und Toxikologie, Universitätsklinikum Bonn	Bonn
Institut für Pharmakologie, Toxikologie und Klinische Pharmazie	Braunschweig
Institut für Klinische Pharmakologie, Klinikum Bremen-Mitte GmbH	Bremen
Institut für Pharmakologie und Klinische Pharmakologie, Universität Düsseldorf	Düsseldorf
Pharmakologisches Institut für Naturwissenschaftler, Universität Frankfurt	Frankfurt am Main
Institut für Pharmakologie, Universität Greifswald	Greifswald
Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Universität Hamburg	Hamburg
Fraunhofer-Institut für Toxikologie und Experimentelle Medizin	Hannover
Institut für Pharmakologie und Toxikologie, Universitätsklinikum Jena	Jena
Institut für Toxikologie und Pharmakologie für Naturwissenschaftler, Universitätsklinikum Schleswig-Holstein	Kiel
Rudolf-Boehm Institut für Pharmakologie und Toxikologie, Universität Leipzig	Leipzig
Institut für klinische Pharmakologie, Universität Magdeburg	Magdeburg
Institut für Pharmakologie, Toxikologie und Pharmazie, Tierärztliche Fakultät LMU München	München
Institut für Pharmakologie und Toxikologie, Universität Münster	Münster
Institut für Klinische Pharmakologie	Stuttgart
Institut für Pharmakologie und Toxikologie, Universität Tübingen	Tübingen
Abt. für Pharmakologie im Klinikum Wuppertal	Wuppertal
Lehrstuhl für Toxikologie der Universität Würzburg	Würzburg

<i>Biochemie</i>	
Institut für Biochemie, Universität Leipzig	Leipzig
Institut für Biochemie und Biotechnologie, Universität Halle-Wittenberg	Halle/Saale
Institut für Biochemie, Universität Erlangen/Nürnberg	Erlangen
Institut für Biochemie, Universität Frankfurt	Frankfurt
Institut für Biochemie und Biotechnologie der Pflanzen, Universität Münster	Münster
Max Planck Institute of Biochemistry	Martinsried
Molekular- und Zellbiologie	
AG Molekularphysiologie Klinik für Anaesthesiologie und Operative Medizin	Berlin
Klinik für Dermatologie der Charité Universitätsmedizin	Berlin
Medizinische Klinik, Abt. Nephrologie, Charité Berlin	Berlin
Institut für Zelluläre und Molekulare Botanik - Zellbiologie der Pflanzen, Universität Bonn	Bonn
MPI für Molekulare Physiologie	Dortmund
Dermatologie, Universitätshautklinik	Freiburg
Biozentrum Klein Flottbek /Molekularbiologie	Hamburg
Institut für Molekular- und Zellphysiologie, MH Hannover	Hannover
FZL, Klinik für Innere Medizin I, Forschungslabor AG Molekulare Kardiologie und Stammzellenforschung	Jena
Institut für Biologie und Umweltwissenschaften, Universität Oldenburg	Oldenburg
Institut für Zelltechnologie e.V	Rostock
<i>Mikrobiologie</i>	
ZIEL Abteilung Mikrobiologie, TU München	Freising
Institut für Medizinische Mikrobiologie, Infektions- und Seuchenmedizin, LMU München	München
Lehrstuhl für angewandte Mikrobiologie, TU Dresden	Dresden
Institut für Mikrobiologie, Universität Düsseldorf	Düsseldorf

UKE Institut für Medizinische Mikrobiologie, Virologie und Hygiene	Hamburg
UFZ-Umweltforschungszentrum Leipzig-Halle GmbH Department Umweltmikrobiologie	Leipzig
<i>Medizinische Forschung allgemein</i>	
Wissenschaftliches LaborMed. Universitäts-Poliklinik	Bonn
Institut für Lasermedizin	Düsseldorf
Martin-Luther-Universität Halle-Wittenberg / Zentrum für Medizinische Grundlagenforschung	Halle (Saale)
Fakultät für Biowissenschaften	Heidelberg
Max Planck Institut für medizinische Forschung	Heidelberg
Institut für Pharmazie, Universität	Leipzig

## 8.4 Fragebogen für Delphi I

Item	Absolutely necessary	Moderately necessary	Neutral	Moderately unnecessary	Absolutely unnecessary
Methodology					
Is the experimental model (growth model toxicological model) explicitly named?	O	O	O	O	O
Is the experimental model adequate for answering the questions?	O	O	O	O	O
Is the variance within each of the examined populations equal? (assumption of homogeneity)	O	O	O	O	O
Are the objectives of the study clearly stated?	O	O	O	O	O
Are the main parameters clearly defined?	O	O	O	O	O
Are the main parameters adequate?	O	O	O	O	O
Is the main outcome a priori defined?	O	O	O	O	O
Is the background of the study clearly described?	O	O	O	O	O
Is the schedule of the study clearly defined?	O	O	O	O	O
Is all relevant literature or previous work appreciated adequately?	O	O	O	O	O
Is the statistical design clearly described (e.g. randomised blocked design, split block design...)?	O	O	O	O	O

Item	Absolutely necessary	Moderately necessary	Neutral	Moderately unnecessary	Absolutely unnecessary
Are the statistical tests appropriate ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the statistical tests adjusted for multiplicity?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are measures of variability given?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are mean effects calculated?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the units allocated to the treatment by a random process?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is treatment allocation concealed?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the conductors of the study blinded?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are all materials (devices, substances) given in detail?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is it clear how many units are treated?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is it clear how many experiments have been done in one setting?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the results internally replicated?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the results externally replicated?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is the observation time long enough?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are all treated units analysed? (intention to treat)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is it clearly stated how many units were excluded from the analyses?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is it clearly stated why many units were excluded from the analyses?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Item	Absolutely necessary	Moderately necessary	Neutral	Moderately unnecessary	Absolutely unnecessary
Test substance (homeopathic)					
Is it clear whether the probes were potentized and/or diluted?	O	O	O	O	O
Is the method of potentization explained (Number of strokes, Korsakov-method,...)?	O	O	O	O	O
Were commercially available substances or self-made substances used?	O	O	O	O	O
Is it clearly described under which conditions and for how long the substances were stored?	O	O	O	O	O
Is information given as to how the authors avoided contaminations of the probes?	O	O	O	O	O
Are dates available (e.g. mother tincture) to allow the calculation of the molar concentration of the test substance?	O	O	O	O	O
Is the exact dose of the treatment given?	O	O	O	O	O
Is information about the form (globuli, tablets) of Application given?	O	O	O	O	O
Is information about the time of Application given? (morning, evening, every two hours)	O	O	O	O	O
Is information about application method given? (oral, injection,...)	O	O	O	O	O

Item	Absolutely necessary	Moderately necessary	Neutral	Moderately unnecessary	Absolutely unnecessary
Controls					
Are the control substances prepared in the same manner as the test substance?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are the control substances prepared with the same liquid as the test substance?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is a control done for each potency?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is it possible to compare each result gained in the probe with that of the identically prepared control/placebo?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is a systematic negative control conducted?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is a positive control conducted?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Experimental setup					
Are the laboratory conditions standardised?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are all biological conditions mentioned? (soil, time of cultivation, time of the year, harvest,...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are all physical conditions described and defined? (temperature, humidity,...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Item	Absolutely necessary	Moderately necessary	Neutral	Moderately unnecessary	Absolutely unnecessary
Is the Charge number of used substances or instruments labelled?	O	O	O	O	O
Are the correct name and species of the used organisms given?	O	O	O	O	O
Is it clear which sex the animals in the study are?	O	O	O	O	O
Are details given concerning the nutrition or incubation of the probes?	O	O	O	O	O
Was an adaptation period of the system planned? (e.g. animal, plant)	O	O	O	O	O
Are Intervention details given? (time, interval, volume,...)	O	O	O	O	O
<b>Results</b>					
Are the results clearly presented in tables?	O	O	O	O	O
Are the results clearly presented in figures?	O	O	O	O	O
Are the conclusions appropriate to data?	O	O	O	O	O
Are the main results pointed out?	O	O	O	O	O
<b>Discussion</b>					
Are the main shortcomings discussed?	O	O	O	O	O
Does the discussion address the main literature in the field?	O	O	O	O	O
Others	O	O	O	O	O



## 8.5 REHBaR: Erläuterungskatalog mit Beispielen

### (1) Title

Explanation: An accurate title is important to indicate what the publication is about and to catch the interest of the reader who is concerned with this research. Make sure that the title includes either the exact name or a description of the used test system as well as the type of intervention. If indicated, terms for identifying the experiments as “in vivo” or “in vitro” should be given.

Activation of mononuclear bone marrow cells treated in vitro with a complex homeopathic medication(Cesar et al. 2008)

### (2) Abstract

The abstract provides more information on what was done in the experiment(s). Give enough details to enable the reader to decide, whether the publication is worth reading or not. To provide a sufficient overview we recommend – as a lot of journals demand already – a short text structured by the following headings: Background, Objective, Materials and Methods, Results and Conclusions. The investigated test system should be clearly defined by a commonly used term or a brief description. Abstracts should inform the reader of the main intention of the experiments, the methods and the intervention. Give the name of the applied substance or other interventions (e.g. heat, radiation) as well as the method of application (e.g. oral, intravenously), the dose (e.g. ml,  $\mu\text{mol}$ ) and the potencies of homeopathic remedies used (potency level and ratio, e.g. 6c, 30x). The abstract accentuates the main results supported by numbers and estimates of associations, variability and uncertainty.

**BACKGROUND:** Efficacy of higher homeopathic potencies is controversial. Universally accepted specific detection assays for homeopathic dilutions do not exist. Basic research has to develop a spectrum of standardized tools to investigate the mode of action and nature of homeopathic potencies. **OBJECTIVE:** Can the shoot growth reaction of dwarf peas (gibberellin- deficient mutants) be regarded as evidence of treatment with homeopathic potencies of plant growth substances? **MATERIALS AND METHODS:** Pea seed (*Pisum sativum* L. cv. Fruher Zwerg) is immersed for 24 hours in homeopathic potency or control solutions for soaking. Plants germinate and grow in a standard cultivation substrate under controlled environmental conditions. Shoot length is measured 14 days after planting. **RESULTS:** A screening of homeopathic potencies (12x-30x) of four different plant growth substances revealed biological activity of certain potency levels of gibberellin and kinetin ( $p < 0.05$ ). Growth stimulation through gibberellin 17x ( $5 \times 10^{-18}$  M) was assessed in six independent replications; results confirmed those of the screening ( $p < 0.05$ ). The effect of gibberellin 17x seemed to weaken during the course of the experiments. **CONCLUSION:** The results back the hypothesis that homeopathic potencies of plant growth substances affect pea shoot growth. Dwarf peas might thus be an interesting system model for studying the action of homeopathic potencies. Further work is required to identify all boundary conditions modulating the reactivity of this system. (Baumgartner et al. 2004)

## **(3–4) Introduction**

### **(3) Background**

The background section should give readers a general idea of the specific topic the publication deals with. Cite pertinent scientific literature or previous work on the subject to document the relevance of the examined topic. You should provide sufficient information on what was going on before this particular research question came up, why the experimental model and the main parameters were chosen and also why they are adequate for answering the research question. Concerning homeopathy, refer to the homeopathic principle which forms the basis of the investigation: e.g. the similia principle, potentisation, proving. The type of homeopathy investigated should be defined e.g. isopathy, classical vs. complex homeopathy.

...Published results of homeopathic studies in poultry are rare. Vizzani and Novelli (1992) examined the ability of homeopathic remedies as growth promotor in broilers and found an effect similar or slightly better than that of the standard antimicrobial growth promotor. However, homeopathic preparations and antimicrobial food additives were not specified, nor was statistical analysis performed... The present study was intended to determine the efficacy of several combinations of isopathic and pluralist homeopathic medicines on experimentally induced colibacillosis in broilers. (Velkers et al. 2005)

### **(4) Objectives/Hypotheses**

Beside the general scientific background, we recommend an illustration of the objectives of the present study. Be aware of what kind of experiments is being presented. For confirmatory experiments, specific hypotheses and clearly defined primary outcome measures are essential. For exploratory experiments, it is often not possible to define primary outcome measures or parameters. In this case you should emphasize the hypotheses inducing the investigations.

On the basis of the experimental evidences in wheat and tobacco models, our hypothesis is that a systematic reduction of variability might be one of the peculiar actions of UHD (Ultra High Dilutions). Therefore we propose to consider variability as a central theoretical issue worthy of study in its own right. (Nani et al. 2007)

## **(5–17) Materials and Methods**

This part of the publication is a detailed list of instructions and information that should enable any interested scientist to replicate the work done. In case a journal cannot provide enough space to publish all details, we recommend making additional information available by using an online appendix or by giving a link to the website of the authors' institute.

### **(5) Materials**

A detailed description of the biological or physical or biochemical model in the experiment is essential. State the name of the used model explicitly: e.g. growth model, toxicological model.

Transparency of the different laboratory procedures is necessary to attest reliability and validity of results. Remember to describe all devices, dilutions and materials as well as all instruments, that were used and to give the proper names, device and lot numbers as well as serial numbers. If the measurement procedure is not commonly used or known, it needs to be specified. To provide an exact description on how many units per experiment were investigated and absolute number of experiments, we recommend a clear outline.

... Luminescence readings were made with a luminometer (Lumistox 300, Dr. B. Lange, Düsseldorf, Germany) at  $15.0 \pm 0.2$  °C. Sterile filter tips with a polyethylene filter (800 µl, Brand) were used to pipette the samples. Samples were kept in cylindrical glass vials (diameter 12 mm, Dr. Bruno Lange GmbH, Düsseldorf, Germany) while the measurements were made... Luminescent bacteria were freeze-dried (*V. fischeri*, strain NRRL-B-11177, supplier Dr. B. Lange). All were from one batch and certified according to DIN EN ISO 11348-3. (Brack et al. 2003)

## **(6) Materials (homeopathy specific)**

All manufacturers and manufacturing processes should be given in detail. If possible refer to a Pharmacopeia or other guidelines as a reference (e.g., Korsakovian single-vial-method, Hahnemannian multi-vial-method or continuous fluxion). If an individual method of preparation was chosen, give explanations as to why the preparation steps differ from the commonly used methods and give enough information on how they were done: Number of strokes, time period of shaking, frequency of shaking, horizontal or vertical shaking, shaking against a soft pad or in the air, using a special machine or manual shaking. State which starting point was chosen and which substance (e.g. mother tincture, D1, nosode) Commercially available preparations should be correspondingly named. If substances were stored, report where and for how long.

... Dilutions were made up in 15 ml sterile polystyrene conical tubes provided with caps; the serial dilutions were prepared with an actual histamine concentration four times greater than the intended final concentration, because the protocol subsequently called for a further fourfold dilution in the cell incubation mixture (see below). A  $4.9 \cdot 10^{-2}$  mol/l histamine solution (designated 1C) was made by diluting 500 µl of histamine stock in 4,500 µl of ultrapure water. From that preparation, a series of incremental 1/100 dilutions were prepared by diluting 50 µl of the preceding solution in 4,950 µl of ultrapure water and shaking by vigorous mechanical shaking (7.5 s at 20 strokes/s, vertical amplitude  $7 \pm 2$  mm) using an instrument from MGA Technologies, Lyon, France. The chosen working dilutions/successions were 2C ( $10^{-4}$  mol/l histamine in the final sample assay) as a positive control, and seven ultra high dilutions from 10C ( $10^{-20}$  mol/l histamine in the final sample assay) to 16C ( $10^{-32}$  mol/l histamine in the final sample assay). .... The histamine and water control dilutions/successions were prepared fresh each day just before the experiments, and stored at room temperature and protected from light until use.. (Chirumbolo et al. 2009)

## **(7) Homeopathic Controls**

Precise description of the manufacturing and the Pharmacopoeias of the test and control substances is indispensable. Make it obvious to the reader why the particular control substance or situation was chosen. Is the control adequate to the objective? Some examples:

5. if the expected effect is caused by the succussion process only, unpotentised solvent is the adequate control
6. if the effect expected is to be specifically caused by a potentised substance, a potentised solvent or another potentised substance (at the same potency level) are adequate
7. if experiment investigates the importance of the succussion step of potentised substances, a combination of potentised solvent and diluted mother tincture is adequate
8. if experiment investigates the simile-principle the adequate control is a variety of substances which represent different degrees of similarity with the diseased state.

Example 1:...As a control, a pool of homeopathic potencies of water 15–20c was prepared according to the procedure described above, except for the starting dilution in which cadmium chloride was omitted. The ... effect was calculated with respect to potentized water in order to account for the physicochemical effects of the potentization procedure. ... (Walchli et al. 2006)

Example 2:...Experiments with highwater dilutions were performed using control test samples made up by an identical procedure, with the only difference that the stock starting solution was pure water rather than histamine..... (Chirumbolo et al. 2009)

## **(8) System Performance Controls**

In basic research it is good practice to include positive and negative controls. If done so, state it; if not, give reasons why such a control was not conducted.

1. Positive controls are performed to demonstrate the reactivity of the system. Describe how the units were treated in a way or with a substance, which causes a certain reaction, and give the results.
2. Negative controls serve to demonstrate the non-existence of carry-over-effects, and in some experimental designs they are necessary to characterize the state of the undisturbed experimental system (e.g. human basophils without activation). Give details on the procedure or state performance of this control.
3. Systematic negative control experiments (also called blind runs or non toxicant tests) are conducted to demonstrate the stability of the test system and to exclude false-positive results (artefacts), e.g. due to differences in spatial position (e.g. of plants in a growth chamber, or of cells in a microtiter plate, due to inhomogeneous temperature or light exposition) or in time order (e.g. treating or measuring objects one after the other). Give details on how they were performed.

... Three blind runs were performed in order to assess the stability of the experimental set-up, i.e. the experiment was performed with the same number of plants ( $525 = 25 \text{ plants} \times 21 \text{ trays}$ ) as in an experiment with homeopathic potencies; however, instead of 21 different parameters (19 potency levels and 2 controls) we used 21 times the same parameter (distilled water). None

of these yielded statistically significant differences between the 21 pseudo-parameters (table 4) ... (Baumgartner et al. 2004)

### **(9) Quality control**

The conductor of the experiment has to guarantee the quality of the study. Therefore we recommend a description of all efforts, which were undertaken to enhance the quality of the study. The report ideally gives information on: training of the experimenter, calibration of devices, supervision of co-workers, information about how the experimenters avoided contaminations of the substances and test systems.

...The experiments were carried out independently by Scherer, Suanjak and the Weber/Welles team in the laboratory of the Interuniversity College...

... An external observer who came to the laboratory, the veterinarian M. Wurn, was responsible for the blinding ...

For reasons of laboratory convenience (danger of cross-contamination) we did not use more than one vial per substance. The project was organized by Endler. (Weber et al. 2008)

### **(10) Object of Investigation**

A detailed description of the researched system is necessary. Define the object precisely: animal, gender, organ, organism, cell compartments (e.g. mitochondria, nuclei), part of cell line, origin of cells, plants, parts of plants, fungi, bacteria, viruses, physical system (e.g. spectroscopy), biochemical (e.g. enzymes), chemical system. The authors should give reasons why the particular system was chosen.

The rat hepatoma cell line Reuber H35 was routinely grown at 37 °C in subconfluent monolayers in plastic flasks (Greiner, Frickenhausen, Germany). Standard growth medium consisted of Leibovitz (L15) medium, pH 7.4 (Flow/ICN Laboratories, CA, USA) supplemented with 10% fetal calf serum (Gibco Life Technologies, Alphen a/d Rijn, The Netherlands), and the antibiotics streptomycin sulfate (100 µg/ml) and potassium penicillin G (100 U/ml). (Wiegant et al. 1994)

### **(11) Experimental setup**

In experiments with plants, specify all important biological conditions like soil, time of cultivation, time of the year when experiments took place, harvest time, charge number of seeds. In experiments concerning animals, gender, age, weight, feeding, holding conditions and intake of fluids should be described. When reporting experiments done *in vitro* (e.g. enzymes, physical parameters) take into account any parameters that could have influenced experimental conditions. Report on preparation procedure and origin of material, incubation time and cleaning methods.

Every detail for a standardisation of the physical conditions be stated (e.g. humidity, temperature, light). Information about any particular equipment for running the experiments is needed. If any procedure was described in an earlier publication refer to that report.

A precise schedule of when what was done as well as a step by step description of the experiment should be given, e.g. in a flow chart. Provide an exact description on how many units per experiment were investigated and on the absolute number of experiments.

... one ml of each solution was placed in aluminum test cups of 20-mm diameter and 2-mm depth and frozen to 253 K (−20°C). Each test cup had a number and the numbers corresponding with each of the four solutions were noted. The frozen samples were kept 1 day at 253 K to achieve stability of the crystalization pattern... Irradiation of the two holders with cups was performed with the Co-60 source Gammacell 220 Excell (GC220E) from MDS Nordion (Fleurus, Belgium). A special Dewar vessel in a size to fit in the radiation chamber was used. The dose selected was 1 kGy. The average dose rate was approximately 30 cGy/second; irradiation time was 58 minutes... Thermoluminescence equipment (IMD electronique, Montpellier, France) was equipped with a Statop 4849 temperature Controller (Chauvin Arnoux. Vaulx en Velin, France). For each measurement, the Dewar of the thermoluminescence equipment was filled with liquid nitrogen. When the aluminium temperature block had reached 78 K, a cup was transferred rapidly to the block and heating was started. The average time to increase the temperature of the block was approximately 20 minutes before the sensors registered a linear increase of temperature in time. Temperature increase was 6°C per minute from 85 K to 235 K. (Wijk et al. 2006)

## **(12) Replication**

State explicitly if experiments were internally or externally repeated and give number of independent experiments (an internal repetition means that the same experiment is replicated with new material and new samples within the same laboratory). Mention if some of the substances or materials were reused.

Example 1: Growth stimulation of through gibberelin 17x ( $5 \times 10^{-18}$  M) was assessed in six independent replications...(Baumgartner et al. 2004)

Example 2: The aim of the present study was to reproduce the original experiment in order to verify whether the same significant results could be obtained working in a different place and with a different experimental team. (Brizzi et al. 2005)

## **(13) Parameters**

All measurement parameters should be clearly defined. Explain why these particular parameters were chosen and why the selected parameters are adequate to answer the research question. In confirmative experiments, declare the principal outcome measure.

Systematic measurements of the specific conductivity were performed on the SDA, using a conductometer, YSI mod. 3200, employing a conductivity cell with constant equal to  $1.0 \text{ cm}^{-1}$ . Before measuring the conductivity of the sample, the cell had to be calibrated by determining the cell constant  $K (\text{cm}^{-1})$ . The specific conductivity,  $\chi (\mu\text{S cm}^{-1})$ , was then given by the product of the cell constant and the conductivity of the solution. For a chosen conductivity measuring cell, the cell constant was determined by measuring the conductivity of a KCl solution with a

specific conductivity that was known with great accuracy for several concentrations and temperatures. All the values of conductivity were temperature corrected to 25 °C, using a prestored temperature compensation for pure water. (Elia 2007)

#### (14) Intervention

We recommend the specification of all drug interventions (homeopathic and/or substantial) as follows:

5. Dose: which volume in which concentration was administered?
6. Time and intervals: when was it applied for the first time and for how long and how often were test and control substances administered?
7. Application mode: how were test and control substances applied? Globules? Tablets? Liquid?
8. Method of application: which way was chosen for test and control substance? Oral? By injection?

Batches of 30 young male Wistar rats, each weighing approximately 70 g, were given, via oesophageal tube, a single dose of 10 mg/kg arsenious anhydride and a tracer dose of 100  $\mu$  Ci/kg  $^{73}\text{As}$  in the form of arsenious acid (Amersham). These were suspended in 5% gum Arabic syrup, and administered in a volume of 0.5ml per 20 g body weight. Twelve hours later oral dosing ( $T_0 + 12\text{h}$ ) the animals were given a single intraperitoneal injection of 1 ml of the Hahnemannian dilution of arsenicum album or the same volume of control (succussed distilled water). Following this the rats were isolated in metabolism cages and allowed water and normal feed ad libitum. (Cazin et al. 1987)

#### (15) Allocation

Describe exactly how it was decided which unit received the test or the control treatment. If allocation was done randomly, describe exactly who generated the randomisation list and which random procedures were used (drawing lots, software etc.).

... Therefore, it was decided to perform a further series of experiments in which  $\text{HgCl}_2$  potencies and controls were tested blind after randomisation by a statistician (RL) and coding by the team supervisor (SH) (experiment 2: see Table 2)...(Witt et al. 2006)

#### (16) Blinding

Give details if and how it was ensured that the conductor was unaware which unit received which treatment (concealment of the randomisation list).

... The experiments were carried out independently by Scherer, Suanjak and the Weber/Welles team in the laboratory of the Interuniversity College. All experiments including application of test and control substances  $10^{-30}$  as well as scoring of the stage of the animals, were performed blind. An external observer who came to the laboratory, the veterinarian M. Wurm, was responsible for the blinding procedures. The same blinding method was used in each case. Substances used for treatment (see below) were prepared in sets each consisting of the test solutions and the

control solution. The plaintext labels were the removed by the person responsible for blinding and replaced with labels bearing encoded designations. The code was not made known until after the presentation of the results ... (Weber et al. 2008)

## **(17) Statistical Methods**

It is essential to give an exact description which statistical methods and tests were applied. Provide an explanation as to why these methods were used and considered appropriate. If appropriate, describe whether or not the statistical tests were adjusted for multiplicity. Furthermore, describe any statistical methods used to control for confounding, and how drop-outs and missing data were handled.

Statistical Analysis: All data analysis was performed with the statistics software 'Statistica 4.1 for Mac' (Statsoft, Inc., Tulsa, OK 74104, USA). If not otherwise stated, p-values refer to analysis of variance F tests. Planned comparisons were evaluated with the LSD (Least Significant difference) test only if the preceding F test was significant ( $p < 0.05$ ). This procedure (protected Fisher's LSD) gives a good safeguard against type I error without being too conservative, i.e. it also gives good security against type II errors... As a complementary statistical analysis, non-parametrical Wilcoxon tests were calculated whenever applicable. (Baumgartner et al. 2004)

## **(18–20) Results**

### **(18) Numbers analysed**

Report every step of calculating and including or excluding results. How were results calculated, how much data contributed to this analysis, and how were missing samples, failures of devices and/or drop outs taken into account?

Depending on the model system it may be imperative to exclude samples (e.g. because of errors in treatment, handling or measurement). On the other hand, it can be important to include drop-outs in the statistical analysis or to otherwise discuss drop-outs (e.g. animals died in an experimental study where survival is no explicit endpoint).

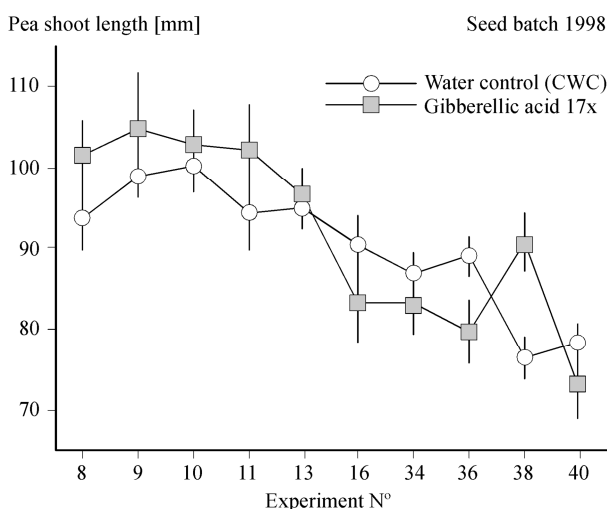
...The entire presented study originally involved 2700 beakers (100 beakers x 27 experiments). Data of 17 beakers were excluded due to spilling. Furthermore, the open controls (n=135, 5 beakers x 27 experiments) were removed from the data set (they had to be included only due to requirements of the image processing software). For the remaining 2548 beakers, images of three time points (day 0, 3 and 7, corresponding to 7644 beaker images) were used for calculation of n= 2548 growth rates for each  $r_{(\text{area})}$  day 0-7, 0-3 and 3-7, respectively. Thus 7644 growth rates entered the statistical analysis ... (Scherr et al. 2009)

### **(19) Data (descriptive)**

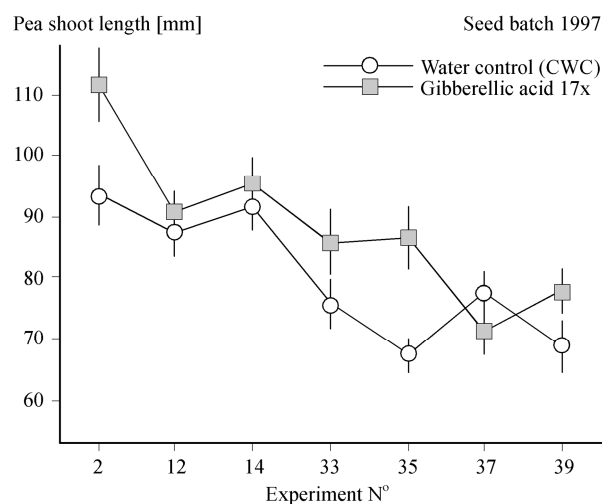
We recommend providing a detailed descriptive summing up of the findings. Exact numbers are best reported in tables. Figures are particularly useful to provide an intuitive summary of the main results. Usually means and standard deviations, or medians and ranges are appropriate if



parameters are scaled continuously. If scaled discretely, it is helpful not only to report absolute frequencies but also percentages.



**Figure 1** Pea shoot length (mean  $\pm$  standard error [mm]) of all experiments involving seed batch 1997, treated with either water or gibberellic acid 17x. Data of the water control group (combined water control, CWC) were pooled from the plants treated with either unsuccussed or succussed water. Data of the experiments No. 2, 12, and 14 have been published earlier.



**Figure 2** Pea shoot length (mean  $\pm$  standard error [mm]) of all experiments with seed batch 1998, treated with either water or gibberellic acid 17x. Data of the water control group (combined water control, CWC) were pooled from the plants treated with either unsuccussed or succussed water. Data of the experiments No. 8–11, 13, and 16 have already been published earlier.

**Table 4** Assessment of variability *within* experiments

Seed batch	Combined water control (CWC)		GA <sub>3</sub> 17x	
	S.D. <sup>†</sup> [%]	n <sup>*</sup>	S.D. <sup>†</sup> [%]	n <sup>*</sup>
1997	35.34	360	32.43	221
1998	25.92	608	28.57	326
1999	20.65	894	20.28	469
2000	22.17	780	23.30	391

Calculation: see 'Data evaluation and statistical analysis.'

Included experiments: see Table 3 (analysis 2).

<sup>†</sup> S.D. = standard deviation [%].

<sup>\*</sup> n = number of plants measured.

Table and figure out of (Baumgartner et al. 2008)

## (20) Data (inferential)

The final step of explaining and presenting the results should be achieved by giving the measures of effect size, i.e. the mean differences between treatment and control. This information is only reliable and valuable if numbers of uncertainty of measurement (e.g. standard error or confidence intervals) and calculation of probability (p-values) are stated as well. Calculation of effect sizes is recommended. Report any p-value calculated, regardless whether the test

was significant or not. Report the p-values of the main outcome not only in tables or figures, but also in the text body. If applicable, report power calculations.

....For a final analysis, data of all experiments with gibberellin 17x were pooled .... A two-way analysis of variance of the dependent variable pea shoot length and of the independent variables treatment (gibberellin 17x and water) and experiment number yielded significant differences for both main effects treatment ( $p = 0.012$ ) and experiment number ( $p = 0.0001$ ), but no significant effect for the interaction ( $p = 0.337$ )... Thus analysis of variance comes to the conclusion that the effect of treatment with gibberellin 17x is fairly reproducible and independent of external factors. On average, treatment with gibberellin 17x increased pea shoot length by  $+(4.6 \pm 1.8)\%$  (mean  $\pm$  standard error) relative to the water control plants. As an alternative statistical analysis, a Wilcoxon Matched Pairs Test was performed for the gibberellin 17x and water treatment means of all 11 experiments .... This test also yielded significant results ( $p = 0.033$ )... (Baumgartner et al. 2004)

## **(21–23) Discussion**

### **(21) Interpretation**

The interpretation should be appropriate to the data presented in the section above. The analyses of results ought to take into account the underlying hypotheses of the study and/or the expected results. It should be linked with existing findings. Shortcomings, like potential bias, e.g. setting or handling, should be discussed. Discuss dangers associated with multiplicity of analyses and outcomes.

... We found for all samples potentized with Aqua bidest., in almost all frequencies where potencies differed from plain solvent, that remedy values were higher than controls values, mostly with significant differences ( $p < 0.01$ )... Surprisingly (and not known from homeopathic literature), remedy-control differences increased with sample age and conductivity, while depending on glass as Container material and Aqua bidest. as solvent. Vessel-induced changes in trace elements may be a necessary co-factor in potentizing and may be responsible for the difference between potentized and plain Aqua bidest.. ... Because of the blinded preparation we consider differences in potentizing (stroke force or frequency) an unlikely explanation. Given these, solitary remedy-control differences being adjacent to differences for more than one remedy are very unlikely to identify a remedy, therefore we did not calculate differences between experiments. (Witt et al. 2005)

### **(22) Evidence**

The author should give an overview of the relevant literature. How are the results to be seen in context with previous research done in the field? Is there an overall validity of data universal for other experiments? Are the limitations of the experiment presented? A general interpretation should demonstrate that all relevant problems and limitations are considered.

...In this paper we show that various chemical stressors at concentrations that do not exert any effect on tolerance development or hsp synthesis in control cells are able to stimulate these parameters when applied to heat-treated cells. In previous work we showed that under specific circumstances a low dose of a stressor is able to exert a stimulatory action in cells that had been previously treated with a high dose of the same agent... As an extension of this work, the pre-

sent paper aimed at identifying the specificity of this paradoxical stimulatory effect exerted by minute amounts of various chemical Stressors. We determined whether the pattern of hsps which is observed upon low dose application in heat shock sensitised cells resembles the pattern of hsps of the first Stressor (heat shock) or of the second chemical Stressor. Interestingly, a significant correlation was observed with the pattern of hsps induced by the second Stressor ( $r=0.621$ ;  $P<0.001$ ) and not with the pattern induced by the first Stressor. In this respect, Cabin and Buchman suggested... The data presented in this paper are partly in agreement with the phenomenon indicated as 'hormesis'; the paradoxical stimulatory action of low doses of conditions that are toxic in higher concentrations. However, there are also differences. Hormetic effects are usually... In this respect, our data are more in agreement with the so-called 'principle of similarity', which suggests that recovery processes in a disordered condition can be stimulated specifically by low doses of compounds that in high doses are able to induce a similar disorder. With respect to the underlying mechanism of the observed stimulation by low dose conditions, further research is required... (Wiegant et al. 1999)

### **(23) Experimental Model**

One of the limiting factors could be an inadequate model or insufficient parameters chosen for proving the hypotheses. When concerned with the specialities of basic research, it is important to demonstrate the eligibility of the experimental model. In experiments investigating homeopathic aspects it is important to focus on the underlying homeopathic theory. Is the expected effect due to the “similia principle” or does it belong to the field of isopathy? We recommend giving all reasons as to why this cell line or plant was chosen and which parameters within this study object are true for the investigation.

Example 1: ... We did not observe any effect of water succussion on pea shoot growth. Thus the present pea model does not seem to sensitive to the unspecific effects associated with succussion, such as air suspension and dissolution, pH alterations radical formation, and enhanced ion release from potentization vessel walls. This further validates the stability of the system, and the reliability and specificity of the effects of gibberellic acid 17x. ... (Baumgartner et al. 2008)

Example 2: ... The authors hypothesize that the stimulatory effect observed is the result of a real homeopathic effect according to the *similia* principle.... (Walchli et al. 2006)

## 8.6 REHBaR – Katalog zur Bewertung der Güte einer Publikation

Title (short form): _____ BR Nr.: _____					
Authors: _____ Journal: _____ Year: _____ Volume: _____					
Date of extraction: _____ Reviewer: _____					
Overall estimation: <input type="checkbox"/> include this paper for a review process <input type="checkbox"/> exclude this paper					
Item	Nr.	Descriptor	Given	page	n.a.
Title	1	Title indicates the experimental model AND the intervention.			
Abstract	2	a) abstract provides an informative and balanced summary of what was done			
		b) abstract provides an informative and balanced summary of what was found.			
<b>Introduction</b>					
Background	3	a) pertinent scientific literature and / or previous work on the subject is cited and gives a sufficient scientific background on the necessity of the presented research			
		b) experimental model is presented			
		c) necessity of the investigation is explained			
		d) name of underlying homeopathic principle(s) (e.g. similia principle, potentisation, proving) and type of homeopathy (isopathy, classical vs complex homeopathy) is given			
Objectives/Hypotheses	4	Objectives and hypotheses with outcome measures:			
		a) for confirmatory experiments: specific hypotheses and clearly defined primary outcome measure is given			
		b) for exploratory experiments: hypotheses inducing the investigations are given.			
<b>Materials and Methods</b>					
Materials	5	Detailed description of all used materials			
		a) devices (with ID),			
		b) substances (non homeopathic) with charge-number (lot-number),			
		c) instruments with sufficient information on origin (factory name, serial-number, town, country.)			
-Materials (homeopathy specific)	6	a) manufacturer, Pharmacopoeia or production method for medications. If degree of potentization is given with D, C, CH or X, preparation method according to Hahnemann is defined.			
		b) potency or steps of dilution,			
		c) dilution method,			

Item	Nr.	Descriptor	Given	page	n.a.
		d) starting point of dilution (e.g. mother tincture, D1, nosode).			
Homeopathic Controls	7	Description of how the control substance was prepared.			
System Performance Controls	8	a) application of a test to demonstrate the stability of the test system (negative control),			
		b) a test to demonstrate the reactivity of the system (positive control).			
Quality control	9	Procedures and efforts used to enhance the quality and reliability of the experimental procedure			
Object of investigation	10	Description for the particular system used in vivo, in vitro, biological, physical, biochemical, ....			
Experimental Setup	11	a) detailed description of experimental conditions and procedures enabling the reader to replicate the experiment(s)			
		b) number of treated units			
Replication	12	In case of internal replications: which materials have been reused; which have been changed.			
Parameters	13	Each parameter measured is named explicitly			
Intervention	14	Details on how and when substances or interventions for each group are actually administered.			
Allocation	15	Method used to generate the group allocation including details (e.g. randomisation, blocking, stratification) has to be explained (term "randomised" is not sufficient).			
Blinding	16	Description if any procedure(s) or intervention(s) were concealed (term "blinded" is not sufficient).			
Statistical Methods	17	a) List of specific tests having been applied			
		b) reason why the test(s) were applied			
		c) information whether statistical tests were adjusted for multiplicity			
<b>Results</b>					
Numbers analysed	18	a) Number of experiments with number of treated units (is this Number identical to Item 11b?)			
		b) information on missing values			
Data (descriptive)	19	a) Results are given in tables or figures (usually means and standard deviations, or medians and ranges are appropriate if parameters are scaled continuously. If scaled discretely, absolute frequencies and percentages are appropriate.)			
Data (inferential)	20	Appropriate measures of effect size (incl. uncertainty and probability) are given (mean differences between treatment and control (standard error or confidence intervals) and calculation of probability (p-values) are stated.)			
<b>Discussion</b>					
Interpretation	21	Interpretation of the results, taking into account study hypotheses or sources of potential bias or imprecision			

Item	Nr.	Descriptor	Given	page	n.a.
Evidence	22	General interpretation of results in the context of current evidence.			
Experimental Model	23	Critical assessment of the chosen model or parameters.			
Reference list	24	The list of cited literature is sufficient.			
Comments:					

EXTRACTION SHEET: State which items are given in the paper. Tick off the column and give page numbers when information is in your opinion sufficient for the particular item. If not or unclear, please leave this column unchecked. If the item is not applicable for the present paper tick off n.a. (not applicable).

## 8.7 Übersicht der bewerteten Publikationen

### ***Publikationen in der Pilotstudie:***

- 1) Bonavida, B.: Overcoming tumor cell drug resistance by low doses of recombinant tumor necrosis factor and drug. In Doutremepuich, C. (ed.): Ultra Low Doses, 1991, S. 27-44 London, Washington: Taylor & Francis
- 2) Bornoroni, C.: Study on the experiments on maize mesocotyl growth by lengthening. In Bornoroni, C. (Hrsg.): Omeomed 92. First International Congress. The Homeopathic Medicine in Europe 1993. Physical-Chemical-Biological and Clinical Research, 1992, S. 117-124 Urbino: Editrice Compositori
- 3) Cristea, A.: Experimental pharmacological Researches concerning the action of Chamomilla homeopathic dilutions on central nervous systems. In GIRI (ed.): 7th GIRI Meeting, 20-21 November, 1993, Montpellier, France, 1993, S. 23-24 Montpellier: Eigenverlag
- 4) Demangeat, J.L.: Modifications des temps de relaxation RMN a 4Mhz des protons du solvant dans les hautes dilutions salines de silice/lactose. In J Med Nucl Biophys, Vol. 16, 1992, S. 135-145
- 5) Harisch, G.: Untersuchungen zur Wirkung von Ubichinon Injeel und Injee forte mit zellfreien Systemen. In BM, Vol. 26(3), 1997, S. 99-104
- 6) King, G.: Pilotstudie zum Thema: "Wirkung von Homöopathika auf Mikroorganismen". 1988, S. 1-19 Ladenburg
- 7) Scherr, C.: Effects of potentised substances on growth kinetics of *Saccharomyces cerevisia* und *Schizosaccharomyces pombe*. In Forsch Komplementärmed, Vol. 13(5), 2006, S. 298-306

### ***Publikationen in der Hauptstudie:***

*\*(durchgestrichen sind die 17 Publikationen, die von der Evaluation ausgeschlossen wurden)*

1. Aggarwal A, Control of taro bright and corm rot caused by *Phytophthora colocasiae* homeopathic drugs. *Plant Dis Rep* 1992;**8**(2): 94-101
2. Anderson D, Statistical analysis of adaptive response in sister chromatid exchanges in human lymphocytes after treatment with very low and extremely low doses of N-methyl-N'-nitro-N-nitrosoguanidine using a study design to control variability. *Brit Hom J* 1999;**88**(1): 7-16
3. Banerjee P, A potentized homeopathic drug, *Arsenicum album* 200, can ameliorate genotoxicity induced by repeated injections of arsenic trioxide in mice. *J Vet Med (A)* 2007;**54**(7): 370-376
4. Baumgartner S, Reproducibility of dwarf pea shoot growth stimulation by homeopathic potencies of gibberellic acid. *Complement Ther Med* 2008;**16**(4): 183-191
5. Belougne-Malfatti E, Combination of two doses of Acetyl Salicylic Acid. *Thrombosis Res* 1998;**90**: 215-221
6. Benveniste J, Transfer of the molecular signal by electronic amplification. *FASEB J* 1994;**8**: A398
7. Binder M, The effects of a 45x potency of *Arsenicum album* on wheat seedling growth. *Forsch Komplementärmed* 2005;**12**(5): 284-291
8. Brack, A., Strube, J.; Stolz, P.; Decker, H.: Effects of ultrahigh dilutions of 3,5-dichlorophenol on the luminescence of the bacterium *Vibrio fischeri*. *Biochim Biophys Acta*, 2003,**1621**: 253-260
9. Carriere V, Importance of temporal factors on the action of infinitesimal doses of interferon alpha beta on the chemiluminescence of mouse peritoneal macrophages. *GIRI (ed.): 2eme Symposium du Groupe International de Recherche sur l'Infinitesimalité, 23-24 Novembre 1988. Monte Carlo: Eigenverlag, 1988, 26*
10. Carriere V, Influence of mouse age on PMA-induced chemiluminescence of peritoneal cells incubated with alpha/beta interferon at very low and moderate doses. *Int J Immunopharm* 1990;**6**: 211-214
11. Datta S, Efficacy of a potentized homeopathic drug (*Arsenicum Album*-30) in reducing toxic effects produced by arsenic trioxide in mice. *Complement Ther Med* 1999;**7**(3): 156-163

12. Demangeat J, Modification of 4MHz N.M.R. Water Proton Relaxation Times in Very High Diluted Aqueous Solutions. *Bastide, M. (ed.): Signals and Images. Selected Papers from the 7th and 8th GIRI Meeting, held in Montpellier, France, Nov. 20-21-1993, and Jerusalem, Israel, Dec. 10-11-1994.* Dordrecht: Kluwer, 1997; 95-110
13. Desai V, Anti-inflammatory activity of Arnica on carragenin induced rat paw oedema. *IAVH (ed.): Proceedings of the 3rd International Congress for Veterinary. Homoeopathy of the International Association for Veterinary Homoeopathy, Münster 4. 6.9.1992.* München: Müller & Steinicke, 1992; 89-113
14. Dittmann J, Temperature dependent influence of As<sub>2</sub>O<sub>3</sub>, HgHPO<sub>4</sub> and KCl on lysosomal acid phosphatase isolated from rat liver. *Biochem Mol Biol Int* 1994;**34**(2): 361-365
15. Dorfman P, Influence of a pool of high dilutions of sulpiride on the estrus cycle modified by administration of ponderal doses of sulpiride to rats. *GIRI (ed.): 5th GIRI Meeting, 29-30 November, 1991, Maison de la Recherche, Paris.* Paris: Eigenverlag, 1991; 33
16. Endler P, Transmission of hormone information by non-molecular means. *FASEB J* 1994;**8**: A400
17. Franco-Lopes D, Viscum album cytotoxicity in Caco-2 cells (in-vitro) analyzed by Dispersive Raman Spectroscopy. *Int J High Dil Res* 2008;**7**(23): 77-78
18. Gaborit J, Etude pharmacologique de rétention et de la mobilisation de l'arsenic sous l'effet de doses infinitésimales d'arsenic. Lille: 1981: 1-83
19. Garcia G, Evaluation "in vivo" de la réponse immunologique du lapin envers un antigène mycobactérien à doses infra-pharmacologiques. *Homéopath Europ* 1997;**3**: 28-29
20. Glatthaar-Saalmüller B, Antiviral Action of Euphorbium compositum and Its Components. *Forsch Komplementärmed* 2001;**8**(4): 207-212
21. Glatthaar-Saalmüller B, In vitro evaluation of the antiviral effects of the homeopathic preparation Gripp-Heel on selected respiratory viruses. *Can J Phys Pharmacol* 2007;**85**(11): 1084-1090
22. Gupta G, Virus Chemotherapy through Homoeopathic Drugs. *Brit Hom Res Group Communications* 1985;**14**: 30-35
23. Harisch G, In-vivo- und In-vitro-Versuche zur Erforschung der Wirkungsentfaltung von Homöopathika. *Forsch Komplementärmed* 1998;**5**(3): 122-131
24. Heckmann C, Einfluß homöopathisch und nicht homöopathisch hergestellter Thyroxinlösungen auf die Metamorphosegeschwindigkeit beim Grasfrosch (Rana temporaria L.). Eberhard-Karls-Universität Tübingen, Fakultät für Biologie: 1997: 1-80
25. Jones R, Plant Responses to Homoeopathic Remedies. *Brit Hom J* 1981;**70**(3): 120-128
26. Jones R, Comparison of Wheat and Yeast as in vitro Models for Investigating Homoeopathic Medicines. *Brit Hom J* 1983;**72**(3): 143-147
27. Kanui T, Wirkungen von Nux vomica D4, D6, D10, Nux vomica-Homaccord ad us. vet. und Atropinum compositum ad us. vet. auf die intestinale Darmmotorik in vitro. *BTM* 1996;**1996**(1): 5-11
28. King G, Pilotstudie zum Thema: "Wirkung von Homöopathika auf Mikroorganismen". Ladenburg: 1988: 1-19
29. Lasne Y, Mise en évidence de signaux physiques émanant de solutions diluées dynamisées ou "homéopathiques". *Bull M.T.S.* 1985;**2**: o.S.
30. Lorenz I, Sensitive flow cytometric method to test basophil activation influenced by homeopathic histamine dilutions. *Forsch Komplementärmed* 2003;**10**(6): 316-324
31. Maharaj S.M, Examination of homeopathic preparations by Fourier transformation-infrared spectroscopy. *Am J Homeopath Med* 2006;**99**(1): 24-30
32. Metelmann H, Die antivirale Wirkung von Euphorbium compositum S. *BM* 1999;**28**(3): 142-146
33. Möhnle U, Der Einfluß von Apis mellifica, Sulfur und ihren Lösungsmitteln bzw. Trägerstoffen auf Bienengift induzierte Stoffwechseleffekte in der Rattenleber. Tierärztliche Hochschule Hannover (Institut für Physiologische Chemie): 1987: 1-233

\*(durchgestrichen sind die 17 Publikationen, die von der Evaluation ausgeschlossen wurden)



34. Moss V, The action of "low potency" homeopathic remedies on the movement of guinea-pig macrophages and human leucocytes. *Brit Hom J* 1982;**71(2)**: 48-61
35. Oberbaum M, Healing Chronic Wounds Performed on Mouse Ears Using Silica (SiO<sub>2</sub>) AS A Homeopathic Remedy. 1997: 191-200
36. Pathak S, Protective potentials of a potentized homeopathic drug Lycopodium-30, in ameliorating azo dye induced hepatocarcinogenesis in mice. *Mol Cell Biochem* 2006;**285**: 121-131
37. Paula Coelho Cd, Therapeutic and pathogenetic animal models for *Dolichos pruriens*. *Homeopathy* 2006;**95(3)**: 136-143
38. Pedalino C.V, Effect of *Atropa belladonna* and *Echinacea angustifolia* in homeopathic dilution on experimental peritonitis. *Homeopathy* 2004;**93(4)**: 193-198
39. Popp F, Bericht an Bonn – Ergebnisse eines Forschungsauftrages zum Wirksamkeitsnachweis der Homöopathie. Kaiserslautern: VGM, 1986: 1-208
40. Pretorius E, Investigating the ultrastructure of platelets of HIV patients treated with immunoregulator, Canova. *Histol Histopathol* 2009;**24**: 399-405
41. Progetti M, Effets curatifs et preventifs de dilutions homeopathiques de sulfate de cuivre appliquées à des racines de lentilles pré ou post-intoxiquées. *Cahiers de Biothérapie* 1985;**88**: 21-27
42. Rastogi D, Homeopathy in HIV infection. *Brit Hom J* 1999;**88(2)**: 49-57
43. Seligmann I, The anticancer homeopathic composite "Canova Method" is not genotoxic for human lymphocytes *in vitro*. *Gen Molecul Res* 2003;**2(2)**: 223-228
44. Stanciková M, Hemmung der Leukozytenelastase-Aktivität *in vitro* mit Zeel T, Zeel comp. und ihren verschiedenen potenzierten Bestandteilen. *BM* 1999;**28(2)**: 83-84
45. Süß W, Struktur und Dynamik von homöopathischen Hochpotenzen – Resonanz /Dämpfungs-/Entdämpfungs-Messungen (REDEM). Süß, W.G. (Hrsg.): *Homöopathische Arzneimittel – wissenschaftliche Grundlagen für die Herstellung, Qualität und Anwendung (5. Leipziger Symposium, 17.05.2003)*. Stuttgart: Deutscher Apotheker Verlag, 2004, 49-63
46. Sukul N, Differentiation of Homeopathic potencies by experimental catalepsy. *Enviroment Ecol* 1986;**4(4)**: 585-588
47. Sukul N, Haloperidol-induced catalepsy in mice and rats suppressed by orally pre-administered potentized *Agaricus*. *Brit Hom J* 1995;**84(1)**: 6-11
48. Sukul N, Potentized Cina reduced root-knot disease of Cowpeas. *Environment Ecol* 1999;**17(2)**: 269-273
49. Sunila E, Effect of homeopathic medicines on transplanted tumors in mice. *Asian Pacific J Cancer Prev* 2007;**8(3)**: 390-394
50. Turco Nasi A.T, Emploi de biothérapiques dans le traitement de souris infectées par *trypanosoma cruzi* résultats préliminaires. *Ann Hom Franc* 1982;**24(3)**: 53-64

\*(durchgestrichen sind die 17 Publikationen, die von der Evaluation ausgeschlossen wurden)

## 9 Danksagung

Neben dem Beruf solch eine Arbeit anzufertigen ist nicht einfach! Welch ein Glück, dass mich so viele Menschen in den letzten Jahren auf diesem Weg unterstützt haben!

Meinem Doktorvater, Herrn Prof. Dr. Gustav Dobos, danke ich sehr für seine Bereitschaft sich dieses Themas anzunehmen und mich durch genau die richtigen Fragen und Tipps immer wieder auf wichtige Punkte gebracht zu haben.

Frau PD Dr. Susanne Moebus danke ich sehr für Ihre kompetente Beratung in allen Fragen zur Grundlagenforschung. Ihre genaue Durchsicht des ersten Manuskripts war mehr als hilfreich für die Strukturierung und das weitere Schreiben der Arbeit.

Bei der Karl und Veronica Carstens-Stiftung und hier vor allem beim Geschäftsführer, Herrn Dr. Henning Albrecht, danke ich für die logistische, inhaltliche und motivierende Unterstützung. Ohne diese Stütze wären z.B. die Befragung und die Durchführung der Konsensuskonferenzen gar nicht machbar gewesen.

Anja Bürlet, Daniela Hacke und Petra Hüsken sei Dank für die emsige Unterstützung beim Beschaffen von Literatur und Daten. Meinen beiden Gutachern, Dr. Jürgen Clausen und Dr. Otto Weingärtner, schulde ich großen Dank für die akribische Durchsicht der zu bewertenden Publikationen und der herzlichen Bereitschaft dies auch zu tun.

Immer wieder hilfsbereit, kompetent, zuverlässig und weder um Rat noch um Tat verlegen waren PD Dr. Frauke Musial, Dr. Stephan Baumgartner und Herr PD Dr. Klaus Linde. Und ganz viel Unterstützung zu jedem Zeitpunkt der Arbeit habe ich Rainer Lüdtke zu verdanken.

Nicht zuletzt danke ich Dr. Carsten Blindauer für seine tatkräftige Unterstützung bei der Fertigstellung dieser Arbeit.

Ganz herzlichen Dank an alle dafür!

## **10 Lebenslauf**

Der Lebenslauf ist in der Online-Version aus Datenschutzgründen nicht enthalten.